Carlos Oya
Florian Schaefer
Dafni Skalidou
Catherine McCosker
Laurenz Langer

Effects of certification schemes for agricultural production on socio-economic outcomes in low- and middle-income countries

A systematic review

March 2017

Systematic Review 34

Agriculture

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding, producing and synthesising high-quality evidence of what works, for whom, why and at what cost. We believe that better and policy-relevant evidence will help make development more effective and improve people's lives.

3ie systematic reviews

3ie systematic reviews appraise and synthesise the available high-quality evidence on the effectiveness of social and economic development interventions in low- and middle-income countries. These reviews follow scientifically recognised review methods, and are peer-reviewed and quality assured according to internationally accepted standards. 3ie is providing leadership in demonstrating rigorous and innovative review methodologies, such as using theory-based approaches suited to inform policy and programming in the dynamic contexts and challenges of low- and middle-income countries.

About this review

Effects of certification schemes for agricultural production on socio-economic outcomes in low- and middle-income countries: A systematic review, was submitted in partial fulfilment of the requirements of grant SR6.1158 awarded under Systematic Review Window 6. This review is available on the 3ie website. 3ie is publishing this technical report as received from the authors; it has been formatted to 3ie style. 3ie will also publish a summary report of this review, designed for use by decision makers, which is forthcoming.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors are also the sole responsibility of the authors. Comments or queries should be directed to the corresponding author, Carlos Oya, co2@soas.ac.uk

Funding for this systematic review was provided by 3ie's donors, which include UK aid, the Bill & Melinda Gates Foundation, Hewlett Foundation and 16 other 3ie members that provide institutional support.

Suggested citation: Oya, C, Schaefer, F, Skalidou, D, McCosker, C and Langer, L, 2017. *Effects of certification schemes for agricultural production on socio-economic outcomes in low- and middle-income countries: A systematic review. 3ie Systematic Review 34*. London: International Initiative for Impact Evaluation (3ie).

3ie systematic review executive editors: Edoardo Masset and Beryl Leach
Production manager: Angel Kharya
Assistant production manager: Akarsh Gupta
Cover design: John F McGill

Effects of certification schemes for agricultural production on socio-economic outcomes in low- and middle-income countries: A systematic review

Carlos Oya
SOAS, University of London

Florian Schaefer
SOAS, University of London

Dafni Skalidou
School of International Development/ University of East Anglia

Catherine McCosker
SOAS, University of London

Laurenz Langer
Africa Centre for Evidence, University of Johannesburg

3ie Systematic Review 34

February 2017

Plain language summary

What are certification schemes and what do they do?
Certification schemes (CS hereafter) set voluntary standards and monitor their compliance (through independent auditors) to make agricultural production socially sustainable and agricultural trade fairer for producers and workers. Although CS are mainly characterized by the description above, they can also be systems engaging in a wider range of activities in policy, advocacy, capacity building, and building markets and supply chains around the objectives of social sustainability. These standards matter for consumers, especially for ethical and safety-aware consumers. They are expected to contribute to a wide range of socio-economic and environmental outcomes, ultimately improving the wellbeing of farmers and agricultural workers, whether employed by corporate plantations or individual producers. They try to achieve these aims through a combination of standard-setting actions, capacity building and training different types of market interventions, such as guaranteed market outlets, price premium, and credit facilities among others, and the application of adequate labour standards.

Do certification schemes work for the wellbeing of agricultural producers and workers in low and middle income countries?
There is not enough evidence on the effects of CS on a range of intermediate and final socio-economic outcomes for agricultural producers and wage workers. Quantitative results are mixed and generally inconclusive, despite positive effects on prices and incomes. Workers' wages do not seem to benefit from the presence of CS. Context matters substantially for the causal chain between interventions of certification schemes and the wellbeing of producers and workers.

How many studies were found?
Initial searches produced 11,000 entries, but a large proportion were irrelevant for the review questions or contained no primary evidence. After screening for selection criteria, the number was reduced to 43 studies for the analysis of quantitative effects, and 136 studies for the qualitative synthesis of barriers and enablers and other contextual factors. The number of studies that could be included in the synthesis of effects for each outcome was very limited, so the results must be interpreted with caution.

Are the studies reviewed credible?
Generally, the quality of the studies is mixed, with a significant number of studies that are weak on a number of methodological fronts, for both quantitative evaluations and for qualitative studies. Much can be done to improve the quality of study design and especially the standards of reporting on methods and results.

What do the results mean?
The results mean different things depending on the target audience. For the ultimate beneficiaries, farmers and workers, the results are not particularly encouraging and show that there is no guarantee that living standards improve through certification. CS, like many other interventions, need the support of other factors and favourable conditions to leave a positive impact. Some of these conditions depend on deep-

rooted socio-economic factors that are unlikely to be substantially altered by certification in the short to medium run. For CS practitioners and businesses trading with certified products, there are four types of lessons: (a) claims about impact should match what is possibly achievable and verifiable; (b) revise standards and interventions away from multiple standards, towards fewer overlaps between systems and rationalisation of interventions considering their relative effectiveness; (c) more attention to acceptable impact evaluation standards is needed to develop adequate impact evaluations; (d) CS need to develop a deeper understanding of context and perhaps adapt and pre-test the type and range of interventions to implement.

For researchers and evaluators lessons include (a) the consideration of a range of methods for different kinds of research questions and a clear understanding of what kind of design is more appropriate for each question; and (b) a more consistently rigorous approach to reporting on methods and results.

Summary

Background

The rise of voluntary standards and their associated certification for agricultural products is a well-established phenomenon in the contemporary dynamics of agricultural trade. Supply chain management is increasingly influenced by a proliferation of standards, and by the organisations setting and monitoring them over a growing number of products. While the objectives of standards and certification schemes (CS) vary, the focus of this review is on social sustainability standards, which are closely related to ethical trading and to schemes that focus on socio-economic outcomes of participants, essentially agricultural producers (particularly smallholders) and wage workers, whether employed by corporate plantations or individual agricultural producers.

Objectives

This systematic review addresses the extent to which, and under what conditions, CS for agricultural products result in higher levels of socio-economic wellbeing for agricultural producers and workers in low- and middle-income countries (LMICs). The primary review question is:

> What are the effects of certification schemes for sustainable agricultural production, and their associated interventions, in terms of endpoint socio-economic outcomes for household/individual wellbeing in low and middle income countries?

The subsidiary review question is:

> Under what circumstances and why do certification schemes for agricultural commodities have the intended and/or unintended effects? What are the barriers and facilitators to such certification's intended and/or unintended effects?

Search Methods

We systematically searched for available literature from a wide range of sources. Several bibliographical databases were consulted. A very significant amount of time was devoted to a systematic search for relevant items through hand searching in targeted databases and websites, including consultation with relevant stakeholders in the community of standard-setting organisations. In this field the 'grey' literature is very important. Thus, the standard bibliographic databases would not be enough to find all relevant material. Papers in English, French, Spanish, German and Portuguese were considered. The references retrieved for this review are up-to-date as of November 2015. Some key references were added in July 2016 as a result of consultations with the ISEAL Alliance.

Selection Criteria

We included studies that evaluated the effects of CS on socio-economic outcomes for agricultural producers and workers. We defined eligible CS as those based on second (industry-level) or third-party certifications thereby excluding own-company

standards. We examined the main types of interventions usually implemented by CS, organized around four groups: (a) capacity building, (b) market interventions (including price interventions, credit support, guaranteed market outlets, etc.); (c) premium-funded social investments, and (d) labour standards. In most cases CS adopt combinations of these groups of interventions. We included studies that report at least one intermediate or final outcome of interest.

For the effectiveness review, we selected studies that use experimental and quasi-experimental methods, and other studies that demonstrated control for selection bias and sufficient confounders. We selected studies that provided relevant comparisons with non-certified groups. For questions on barriers and facilitators and contextual factors we searched for and screened qualitative studies that reported on relevant outcomes, that had sufficient reporting on methods, and provided substantive evidence on key selected themes to complement the effectiveness review.

We used a combination of single screening with substantial piloting and supervision in initial stages, and double screening with arbitration for disagreements in coding and inclusion/exclusion decisions for full-text review.

Data Collection and Analysis

We developed separate coding tools according to the requirements of our two review questions. To compare effects on variable outcomes across studies we calculated standardised mean differences. The quantitative results were synthesised using inverse variance-weighted random effects meta-analysis. Only one effect size per outcome per study was included in any given synthesis. The analysis of qualitative material was organised around three main thematic areas: barriers and enablers in implementation dynamics; distributional dynamics, including gender equity issues; other internal and external contextual factors and barriers and enablers.

Results

The initial search returned 10,753 studies, which, after dropping duplicates, a large number of irrelevant papers, and applying the selection criteria, were reduced to a final sample of 43 studies from 44 papers for review question 1 (effectiveness), and 136 studies from 114 papers for review question 2. All were published between 1990 and 2016. The majority of our material comes from research reports, working papers, book chapters, and theses.

The included studies for the quantitative and qualitative syntheses provide evidence on a range of rural settings in LMICs, with dominance of cases from Latin America. Despite the fact that there are many CS operating with agricultural commodities, included studies only cover a group among them (12 CS), which have attracted more research in the form of impact evaluations. Fairtrade certification is particularly well represented in the literature, with over half of the total number of included studies. Several agricultural products are covered by the included studies but coffee (38%) and fruits (17%) combined account for more than half of studies. In terms of population, a large majority of studies (77%) focus on agricultural producers, whereas the research on employment outcomes is rather limited.

The quality of the included studies is mixed. The proportion of quantitative studies with high risk of bias ratings was relatively large. There are no randomized controlled trials (RCTs) but there is a range of quasi-experimental designs employing different techniques of data analysis. Given the paucity of calculable effect sizes per outcome and the variety of methods used in different studies the meta-analysis encountered difficulties, and the number of studies with low or moderate risk of bias included for the synthesis of effects for each outcome is very small. Although there are many included qualitative studies of high quality, especially ethnographic research, the overall quality of this group is mixed as well. Several studies, especially non-ethnographic contributions, are only borderline in terms of minimum reporting standards.

In terms of quantitative results, we find that the available quantitative evidence does not give a clear picture of the impact – or lack thereof – of certification schemes. The synthesised effects for our key intermediate and final outcomes are summarised below. For each outcome we present the difference between certified groups and control groups in standardised percentages, with a central estimate and a likely range around the estimate, which reflects the uncertainty inherent in the estimate, added in parentheses.[1]

- *Yields*: We found no clear effect on yields. While certification is associated with a decrease in yields of 20%, the overall effect is not statistically significant (central estimate -20%, range from -52% to 19%; SMD -0.42, 95%-CI from -1.23 to 0.39). The five studies synthesised for this outcome range from negative to positive in their effect sizes. One study was rated as having low risk of bias, and two studies each were rated as moderate and high, respectively.
- *Price*: Prices for certified producers were 14% higher than for non-certified producers (range from 4% to 24%; SMD 0.28, 95%-CI from 0.09 to 0.49). Three of the four studies we synthesised for this outcome provided positive effect sizes. One study was rated has having high risk of bias while the other three were rated as moderate. The overall effect is statistically significant.
- *Income from certified production*: Incomes from the sale of produce were 11% higher if the produce was certified (range from 2% to 20%; SMD 0.22, 95%-CI from 0.03 to 0.41). For this outcome we synthesised ten studies whose individual effect sizes ranged from negative to positive, though none of the negative effect size estimates were statistically significant. Half of the studies were rated as having moderate risk of bias and the other half as high. The overall effect is statistically significant.
- *Wages*: We find that wages for workers engaged in certified production were 13% lower than for workers working uncertified employers (central estimate -13%, range from -22% to -3%; SMD -0.26, 95%-CI from -0.46 to -0.06). Of the eight studies synthesised all but two provide negative effect size

[1] These standardised percentages are statistical constructs that rely on a number of assumptions. They are presented here are only to convey a more intuitive measure of the size of the reported effects. For more information on how these measures are constructed, please see Section 3.3.4.

estimates and the positive effect size estimates are not statistically significant. One of the studies was rated as having low risk of bias, while five were rated as moderate and two as high risk. The overall effect is statistically significant.

- *Total household income*: Effects on the total household income of farmers are unclear. While household incomes of farmers engaged in certified production were 6% higher than those of households not engaged in certified production, the overall effect is not statistically significant (range from -3% to 16%; SMD 0.13, 95%-CI from-0.06 to 0.32). The effect size estimates for individual studies range from negative to positive, though all statistically significant studies provided positive estimates. Four of the studies synthesised were judged to be of moderate risk of bias, while the other four were rated has high risk.
- *Assets/wealth*: We found no statistically significant effect on wealth. Certified producers on average had slightly higher wealth levels than uncertified producer who had been selected to be similar to them, and the overall effect was a 3% increase in assets, but this effect was not statistically distinguishable from zero (range from -7% to 13%; SMD 0.05, 95%-CI from -0.15 to 0.26). For this outcome we had just two studies, both of which provided positive effect sizes. One study was rated has having high risk of bias, the other as moderate.
- *Illness*: We also found no clear effect on producer's health. Pooling the included studies suggests a 7% lower incidence of illness in certified producers compared to non-certified producers, but the overall effect is not statistically significant (central estimate -7%, range from -16% to 2%; SMD -0.15, 95%-CI from -0.32 to 0.03). Please note that, as these findings concern illness, a negative synthesised effect means an improvement in health. Just two studies provided estimates for this outcome, both of which pointed towards a lower incidence of illness. Both studies were rated as having high risk of bias though.
- *Schooling*: children in households of certified producers receive 6% more schooling than children in households of non-certified producers (range from 0% to 12%; SMD 0.12, 95%-CI from 0.01 to 0.24). The individual effect sizes provided by included studies range from negative but not statistically significant to positive. Three of the five studies synthesised for this outcome were rated as having high risk of bias, the other two as moderate. The overall effect is statistically significant.

In most cases, disaggregation by type of CS did not yield conclusive results, although for some CS results were more mixed than for others. Such is the case of Fairtrade for yields and income measures.

The qualitative synthesis discussed a wide array of factors affecting the causal chain in different nodes along the chain, such as: producer organisations (POs) and their characteristics, particularly heterogeneity and power relations within them; relations with buyers and exporters; business models linking buyers and producers (whether open spot markets, contract farming or a mix); national institutions shaping the dynamics of agricultural trade and labour relations; barriers imposed by direct and

indirect certification costs, which negatively affect adoption or the size of benefits accruing to producers; availability of additional external support, often critical for adoption and sustained maintenance of standards; inconsistency in monitoring and auditing practices; heterogeneity of participant groups and the effects of inequality on POs management and the sharing of benefits; difficulties in addressing deep-rooted structures of inequality based on gender; the relative invisibility of large segments of agricultural wage workers, notably those employed by small farmers.

The mixed and inconclusive quantitative effects, combined with the wide range of contextual factors to take into consideration, underline that CS operate in complex environments with multiple interventions, goals, actors and contexts, and as such they do not operate in a social, institutional and economic vacuum.

Authors' Conclusions

Overall, we found mixed results and a dominance of weak or not statistically significant effects. There were both positive and negative effects for different outcomes. Even within a given CS there substantial variation in effects across different outcomes. Thus, it is hard to conclude anything about whether any particular CS performs better compared to others over a range of outcomes. Without more systematic high-quality quantitative evidence on intermediate and final outcomes it is difficult to draw meaningful conclusions with actionable findings. Context hugely matters, as the range of contextual factors and barriers and enablers is vast. This is not surprising and most Theories of Change developed for selected CS acknowledge the centrality of context specificity. Nonetheless, the reviewed qualitative research reveal a number of key barriers and facilitators or contextual features that seem important to understanding the impact of CS. Practitioners can extract some lessons about the kinds of contextual factors that seem prominent in mediating the impact of their interventions, such as the characteristics of POs with which they partner, the deep-rooted social relations of inequality, including gender dynamics, in rural areas of LMICs; the direct and indirect certification costs, and their determinants; the specificities of each supply chain and especially existing relations between established buyers and producers; and the national and local contexts of regulation and economic development.

There are various implications for researchers. First, there is scarcity of high-quality impact evaluations, and a disproportionate attention to some CS and almost no attention to several other CS. The volume of research with rigorous study designs has fortunately expanded in the last 10 years but this review certainly calls for more studies and on more outcomes, especially on employment effects, which have received less attention so far. Second, mixed-methods theory-based evaluations with appropriate counterfactual designs are likely to generate more valuable findings, given the importance of context and the need to link effects with barriers and facilitators in each study. Third, reporting standards must be improved, so published papers should devote more space and attention to reporting details of how research was conducted, limitations and all the relevant statistical information. Many studies had to be excluded from this review or from effect size calculations because of basic reporting gaps.

Contents

List of figures and tables

Abbreviations and acronyms

2SLS	Two-stage least squares
CI	Confidence interval
CS	Certification scheme
DFID	Department for International Development (UK)
DID	Difference-in-difference
FIML	Full information maximum likelihood
FT	Fairtrade
GAP	Good agricultural practice
GVC	Global value chain
ITC	International Trade Centre
IV	Instrumental variable
JB	Joint body
LMICs	Low and middle income countries
MSP-SQ	MPS-socially qualified (MPS is not an acronym)
NGO	Non-governmental organization
NTAE	Non-traditional agricultural exports
Org	Organic certification
PI	Principal investigator
PO / SPO	Producer organization / small producer organization
PSM	Propensity score matching
RA	Rainforest Alliance
RoB	Risk of bias
RSPO	Roundtable for Sustainable Palm Oil
SE	Standard error
SMD	Standardised mean difference
ToC	Theory of change
UoA	Unit of analysis
VSS	Voluntary social standard

1. Background

1.1 The problem, condition or issue

Certification schemes (CS hereafter) that design and monitor standards (generally 'voluntary standards' or VS) for agricultural production are expected to contribute to a variety of aims, such as improving the wellbeing of producers and workers supplying certified products, as well as offering some assurances to consumers that the products they consume are safe (from the use of pesticides, for example) and have been produced according to ethical standards. The claims around impact can be substantial. Some schemes aim to strive to improve lives of workers and make sure 'that the rights of all workers are respected';[2] most work towards an overarching aim such as 'to improve impacts on people and the environment';[3] by proposing alternative trade relations others aim to help consumers 'get farmers a better deal... and that means they can make their own decisions, control their futures and lead the dignified life everyone deserves'.[4] Of course, many of these claims are aspirations about how certification interventions could potentially contribute to improve the wellbeing of producers and workers. CS therefore operate in the realms of both trading relations and of direct support to producers to make trade better for these potential beneficiaries. This background section will provide the context of the issue, particularly the trends in agricultural trade and the rise of standards and certification, which has accelerated since the 1990s. It will then define the nature and scope of interventions through the certification of agricultural commodities and describe how the intervention of CS is supposed to work to improve the wellbeing of agricultural producers and workers, the two key groups of participants and beneficiaries considered in this review. The chapter ends with a discussion of the relevance and importance of this review.

1.1.1 Agricultural trade dynamics and welfare

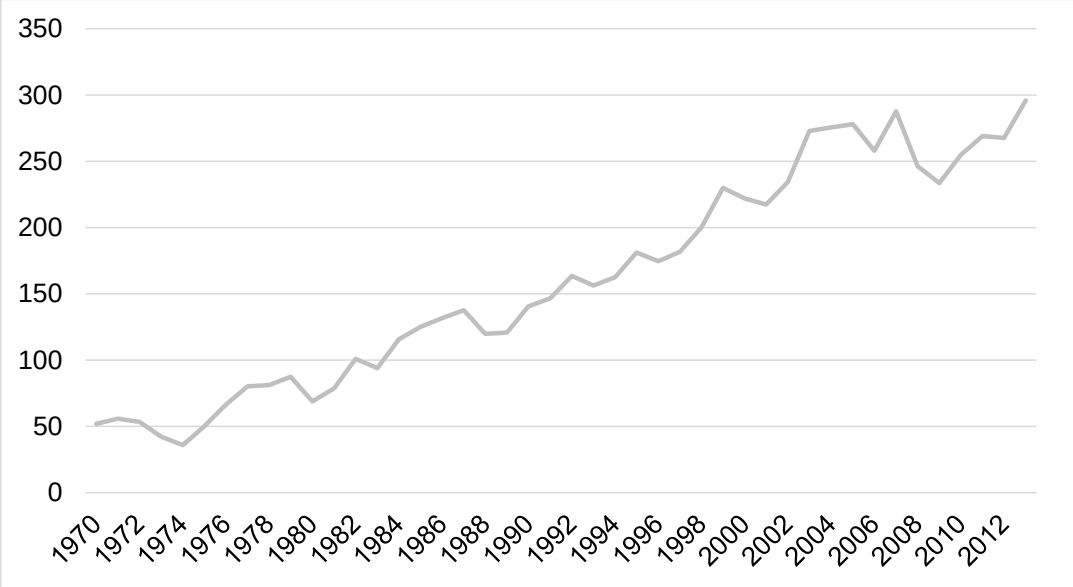
The role of international trade in reducing poverty and increasing welfare in low- and middle-income countries (LMICs) remains an issue of controversy and debate (Winters, 2003; McCulloch et al, 2001). While the share of agriculture in world trade has consistently declined over time, trade in agricultural commodities has expanded continuously in real terms, at least until the global recession following the 2008 financial crash. The volume of agricultural trade has more than doubled since the 1990s (**Figure 1**), as global value chains (GVC) expanded in size and outreach, partly driven by substantial reductions in transport costs, improvements in logistics and various technological developments facilitating the export of perishable goods and giving impetus to the dynamism of non-traditional agricultural exports – NTAEs (Daviron and Gibbon, 2002; Hallam et el., 2004).[5]

---

[2] https://www.utz.org/what-weve-achieved/better-life/
[3] http://www.isealalliance.org/our-work/improving-impacts
[4] http://www.fairtrade.org.uk/en/what-is-fairtrade
[5] Non-traditional agricultural exports include mainly fruits and vegetables, and also specialty produce (e.g. chillies, pepper, paprika etc.) and some processed food (tomato paste, canned mushrooms among many other products). NTAEs are an important group among certified agricultural commodities.

Whether the benefits of agricultural trade expansion and globalisation trickle down to agricultural producers in low- and middle-income countries (LMICs) and especially to the most vulnerable actors in the economy is a contentious issue. Winters (2002) argues that positive effects are likely in the long term, but in the short term trade liberalisation can have adverse effects on the poorest. This is likely to happen to agricultural producers in developing countries, as Nicholls and Opal (2004) highlight, where deficient production and market conditions (poor market information, limited access to markets and credit, lack of ability to adapt rapidly to market changes, among others) are coupled with chronic macroeconomic failures, such as the lack of infrastructure and investment, heavy dependence on only few primary commodities and governance problems. Primary commodity producers are often particularly vulnerable to price volatility and inadequate and asymmetric price transmission mechanisms. Overall, the transmission mechanisms linking global trade in agricultural products with poverty reduction are complex and entail winners and losers, opportunities as well as barriers (Morrison and Murphy 2004).

Figure 1: Agricultural export index in real terms (world exports): 1970-2013



Source: Own elaboration from FAOSTAT data

A voluminous literature on global value chains (GVC) in agriculture has emerged over time and especially since the 1990s, in the wake of structural adjustment programmes and liberalization policies in most LMICs. The complexity of agricultural commodity markets and associated GVCs has been increasing as new technologies, actors and linkages with other markets (notably with finance) emerge, requiring generally more demanding capabilities and performances (Gibbon and Ponte 2005; Ghosh, 2010). Some commodity markets are more dynamic than others because the pessimistic assumption of an exhausted income elasticity of demand only works for certain commodities (CBI 2005). There is also increasing diversification of export destinations outside the EU and the USA for a wider range of agricultural products

(USDA 2006). All these trends may impact on the shifting spectrum of opportunities and constraints faced by agricultural producers and workers.

The incorporation of smallholder producers in agricultural GVCs has been a leading theme in this literature (Vorley et al, 2007; Ouma, 2015; Gibbon and Ponte, 2005; Daviron and Ponte, 2002). Attention has also been paid to how the integration of agricultural producers in GVCs affects the employment conditions of millions of agricultural wage workers (Dolan, 2004; Barrientos and Smith, 2006). GVCs (or global production networks to use another widely known term) operate with an increasing degree of flexibility in terms of supply chain management, sourcing and timing (Gereffi and Korzeniewicz, 1994; Gibbon and Ponte, 2005). A key research question in much of this literature is to what extent the incorporation of producers, especially smallholder farmers, and agricultural wage workers, brings positive effects on their living standards and working conditions (Vorley et al, 2007; Maertens and Swinnen, 2009). Different forms of production coexist for a wide range of agricultural products, partly because of the inherent production characteristics which may be more suitable to large plantation or smallholder production, partly because of historical developments in land property regimes.[6] The variation of forms of production also reflects different historical trajectories across regions, e.g. the greater presence of large-scale plantations in Latin America compared to Sub-Saharan Africa or East Asia, but also within countries, where production conditions may also vary from one product to another or location to another. Overall, smallholder farmers and agricultural wage workers constitute the bulk of the rural poor in most LMICs. Therefore, the impact of GVCs on agriculture in these contexts has important implications for poverty dynamics and these two groups of people dependent on agricultural production. Since exclusion from markets and GVCs is often considered one of the chief causes of poverty in developing countries, a related question is whether 'incorporation' is beneficial or adverse (Hickey and du Toit 2013).

Incorporation into GVCs and agricultural commodity markets comes with different kinds of opportunities and constraints. Competition and entry barriers in agricultural GVCs can be substantial, and are determined by a configuration of multiple factors, but three types of issues seem to play a prominent role across commodities: economies of scale, especially in distribution and marketing; cost advantages, partly related to prime movers and technology factors; product differentiation advantage, itself a function of quality and successful branding (Gibbon and Ponte, 2005, p. 125). Product differentiation is also associated with the fact that international markets for agricultural commodities are increasingly demanding in terms of quality and production conditions, the latter often related to social and environmental sustainability (Henson and Humphrey, 2010).

[6] For example, in some contexts land reforms and the replacement of plantations with cooperatives have moved production from large-scale plantations based on hired labour to smallholder farms more reliant on family labour (Ruben and Fort 2012, on the case of coffee in Peru).

1.1.2 The rise of standards for agricultural production

A growing literature has studied the ways in which buyers have been imposing an ever expanding range of conditions to enter markets in buyer-driven GVCs, precisely as a way of creating product differentiation (Gibbon and Ponte, 2005; Henson and Humphrey, 2010; Maertens and Swinnen 2009). At the same time, ethical trade social movements have since the 1990s also successfully lobbied for 'fair' market conditions to counteract the possible negative effects of 'free' markets on vulnerable producers and workers in LMICs (Barrientos, 2000; Raynolds, 2009; Dragusanu et al, 2014). Many of the CS originating in movements for ethical trade aimed to eliminate middlemen and shorten the chain between producers and consumers, thereby reducing volatility and transaction costs that generally penalised smaller producers (Ruben, 2012; Raynolds and Murray, 2007).[7] The variety of CS and associated requirements on product attributes and production processes, have contributed to creating a range of market segments defined by several product characteristics, sometimes related to intrinsic quality, sometimes linked to ethical and social standards in production processes (Gibbon and Ponte, 2005). When quality and origin matter, for example, demand trends in consuming countries have included the emergence of the 'specialty coffee market', which is becoming ever more important as traceability and other specific features become valued by consumers (Daviron and Ponte, 2005). Similar trends can be observed for products whose demand expands and is increasingly differentiated, as in the case of wine (Ponte 2009). At the same time, some of the emerging standards that impose new demands on agricultural producers have originated in public regulation for food safety. A good example is the range of EU food safety regulations that have converged with expanding private/voluntary standards to define a set of 'good agricultural practices' primarily concerned with consumer safety but increasingly linked to environmental and labour standards upstream. Thus EUREPGAP (which became GlobalGAP in 2007), which was set-up by concerted action from different retail groups, have also contributed to the expansion of spaces of standard setting, and the emergence of an auditing industry required to monitor compliance and evaluate impact.[8] The emergence of these 'standardising networks' responds to contrasting orientations, one market and consumer driven (exemplified by EurepG.A.P./GlobalGAP) and one more concerned with poverty in developing countries and the plight of small producers and workers (Fairtrade and other ethical trade initiatives), and others somewhere in between (Aasprong, 2013). However, even within social movement-driven CS, there is a continuum of vendors (those who ultimately sell certified products) from market-driven corporations looking for a new product niche to 'mission-driven' enterprises 'that advance alternative relational and civic values' (Raynolds and Greenfield, 2015, p. 31). In sum, the factors underpinning an increasingly differentiated and integrated trade in agricultural commodities and the

[7] In a context of increasingly volatile prices, especially after the 2007/08 global food price crisis, CS interventions to give more certainty to producers may have become more significant. See Ghosh (2010) on the linkages between speculation and agricultural commodity markets.

[8] The organisation behind the CS spells it GlobalG.A.P., which stands for global good agricultural practice.

actors involved in setting an ever-growing list of requirements for producers and exporters are multiple and driven by different global forces.

The uptake of sustainability standards keeps increasing, as (social) alternative trade movements and conventional businesses converge towards a set of common goals derived from different logics. The ISEAL Alliance[9], a leading umbrella membership organisation for a range of sustainability standards systems, notes that, as companies see the value of certification in their core business operations, they gradually move 'from a defensive to a proactive position in sustainability…[i.e.] from a reduction of risks to an increase in profits'.[10] This logic applies to conventional private business rather than ethical trade movements. Overall the motivations to expand certification and the use of voluntary social standards (VSS) depend much on the sector, the origins and type of business. For some actors, consumer expectations and the attainment of a sort of 'social licence' are a major motivation. For others, having a set of VSS helps meet other established regulations. Product differentiation and branding, as mentioned above, also play an important role insofar as they have become a key aspect of competition in a global environment. The setting of standards thus creates a set of product specificities that require some form of certification, as in the case of traceability and other product attributes associated with premium quality (e.g. in the case of specialty coffee), or allays consumer concerns about production conditions (as for instance in the case of cocoa or fresh fruits and vegetables). Nowadays there are many sectors where different kinds of standards and especially VSS are applied.[11]

The International Trade Centre, has created a database of sustainability standards, including over 210, across different sectors and a huge range of products. [12] It is not surprising that the geographical reach of standards has expanded considerably since the early 1990s. All major regions of the world are now integrated in a global system of public and private standards affecting an ever growing number of products and services. However, a recent study by the International Trade Centre (ITC) shows that there are still differences in terms of penetration of standards across regions (ITC, 2016). OECD and high-income countries have much higher numbers of standards per country compared to developing countries. This is not surprising given that key drivers of standard setting (consumer demands, product differentiation, etc.) are stronger in high-income countries. Within the developing world and among LMICs, Latin America is the region with the highest number of standards per country at 41, while sub-Saharan Africa has 29, only followed by the Middle East and North Africa where each country has an average of 27.4 standards, applied to a range of products and services. While the number of standards, countries, producers and certified commodities continues to expand, it is very difficult to find an aggregate assessment

[9] ISEAL's mission is to strengthen sustainability standards through a code of practice that all members (as CS) apply in their own standards. See description at http://www.isealalliance.org

[10] http://www.isealalliance.org/online-community/blogs/what-is-the-business-case-for-certification-in-2015

[11] See http://www.isealalliance.org/our-sectors

[12] See http://standardsmap.org/

of the proportion of traded agricultural volume that has some kind of certification for sustainability standards. For some schemes there are estimates of the proportion of world exports that go through certified channels. For example, the Fairtrade product with the highest proportion of world traded output is bananas with approximately 2% in 2014, up from 0.5% in 2004 and 1.7% in 2012.[13] It is likely that a majority of existing standards (excluding public mandatory standards in big markets such as the EU or those applied by all main private supermarkets) concern relatively limited volumes of total traded production, even if the proportions are rising rapidly.

Notwithstanding a relative small market share globally, evidence suggests a rapid growth in sales of certified products in OECD countries in particular. There is more systematic evidence on the case of Fairtrade (FT). For example, during the period 2010-2014 the number of participants in Fairtrade (both direct producers and workers) increased from 1.1 million (of whom almost 1 million were small producers) to 1.65 million (of whom 1.45 million were small producers) with 64% located in Africa and the Middle East (Fairtrade International, 2016). In the UK in 2013 alone 'sales of Fairtrade products exceeded an estimated value of £1.7bn, a 12% increase on 2012' (Fairtrade Foundation, 2014, p. 11). Other similar labels such as Utz have also grown rapidly. A recent annual report estimated a 90% growth in Utz-certified coffee and tea production and a staggering 1,200% for cocoa, adding that 'the market for sustainability is stronger than ever… [with] more than a million farmers and workers … part of the program…[and] 10% of global coffee production now Utz certified' (Utz, 2016, p. 9). In the case of palm oil the share of Roundtable for Sustainable Palm Oil (RSPO) certified production has reached 18% of global palm oil production in 2015.[14]

A different kind of certification scheme, GlobalGAP, emerged in 2007 out of Eurep-GAP, which had been established in 1997 by a group of European retailers to address growing concerns regarding product safety, environmental impact and health in the production of fresh fruits and vegetables, a leading high-value agricultural commodity in the past two decades (GlobalGAP, 2016).[15] GlobalGAP targets mostly larger agribusiness in Europe, but its outreach is quickly expanding in LMICs (GlobalGAP, 2016; Bain, 2010; Masood 2014). In the last ten years this system has grown to the point that it now includes more than 228 certified products, over 160.000 certified producers (from 18,000 in 2004) in more than 118 countries (GlobalGAP, 2016). The proportion of producers who are GlobalGAP certified in 2015 stood at 65% in Europe, 12% in South America and 11% in Africa as the regions with the highest proportions (GlobalGAP, 2015). The system in place includes 136 accredited independent certification bodies that apply the standards

[13] Data from Fairtrade website http://www.fairtrade.net/products/bananas.html, Raynolds and Greenfield (2015, Table 2.2) and world export estimates (in tonnes) from FAOSTAT http://faostat3.fao.org/home/E . Ruben (2016) more recently suggests that 'notwithstanding high annual growth rates, the overall [Fairtrade] share in the world market remains limited to 2-3% in selected commodities'. Based on Raynolds and Greenfield (2015) estimates for green coffee in 2012 the proportion in that case was only 1.1%.
[14] http://www.rspo.org/about/how-we-work
[15] http://www.globalgap.org/uk_en/who-we-are/

and criteria provided by GlobalGAP with some adaptation to national contexts. It is not possible to establish the volume of production certified, but for products in the category of fresh fruits and vegetables it is likely that a significant volume of trade is concerned especially in the EU, where the retail sector imposing such standards concentrates a large proportion of sales (Ouma, 2015; Reardon et al, 2009).

1.1.3 The effectiveness of standards and certification schemes

Since such standards increasingly determine, directly or indirectly (through competition) the terms of integration of agricultural producers in LMICs into global supply chains (Gibbon and Ponte, 2005; Henson and Humphrey, 2010), an important debate has emerged about the effectiveness of certification in raising the welfare of producers – small- and medium-scale farmers and workers in particular. While some standards arise out of concern for food safety and consumer preferences, organisations in networks like ISEAL increasingly acknowledge and pro-actively engage with the idea that certification should also be beneficial for the producers of commodities.[16] As noted in the previous section, the idea of social sustainability and ethical trade arises from the aspiration to break down the distance between parties in a market exchange (consumers and producers) and allow for consumers to play an active role in improving the conditions of producing sourcing the products they consume (Luetchford, 2008; Raynolds and Long, 2007). CS, whether the focus is on social, quality or environmental standards, work towards establishing this new 'connectivity'. On the production side of the relationship, it is expected that producers (or their organisations) take up certification because of potential beneficial effects on prices, access to more remunerative markets, quality, productivity and, as well as of access to continuous support through CS and associated organisations promoting certain types of certification. According to ISEAL consultation forums the current business case for certification among producers now rests more on reducing dependence on demand for certified goods and premium payments, and focuses on mechanisms to increase productivity and quality while reducing certification costs.[17] These discussions show that social sustainability goals are associated with a growing number of possible outcomes for producers of certified commodities.

Given the wide range of CS, certified products and countries involved, it is perhaps not surprising that impact evaluations have differed greatly in their results. Many studies tend to report mixed findings with some positive and other negative elements, or cases where effects are only marginal (Nelson & Martin, 2013, Ruben, 2012). Some have found that CS may actually undermine the incomes of the poorest farmers (Henson and Jaffee, 2008), or that positive effects dissipate due to over-certification (de Janvry et al, 2014); some reported positive impacts for some certification types, but not others (Chiputwa, Spielman and Qaim, 2014), others found effects only for richer farmers (Hansen and Trifkovi□, 2014), while still others showed how CS can help raise rural incomes and reduce poverty (Maertens and

[16] http://www.isealalliance.org/our-sectors .

[17] http://www.isealalliance.org/online-community/blogs/what-is-the-business-case-for-certification-in-2015 . A greater focus on productivity is exemplified by Utz in contrast to Fairtrade as well as in the growth of certification for plantations.

Swinnen, 2009). In the case of fair trade[18] standards, which have been more widely researched, the evidence from primary studies appears inconclusive, mixed and very context-specific, while the quality of studies measuring effectiveness is uncertain, as a number of studies have pointed out (Ruben, 2013; Cramer et al, 2014; Valkila and Nygren, 2009; Terstappen et al, 2013; International Trade Centre, 2011; Nelson and Pound, 2009; Nelson and Martin, 2013). This debate therefore continues and is likely to become increasingly important to policy as the sales of agricultural commodities through market channels that require these kinds of certification expand rapidly.

1.2 The Intervention: description of certification schemes

1.2.1 The types of standards and certification schemes

Standards for agricultural products have different origins and multiple aims, as suggested in the previous section. They may be public mandatory, as in many EU food safety regulations (phytosanitary rules), public voluntary, private mandatory and private voluntary (Henson and Humphrey, 2010). They may be developed by companies, industry associations, governments or NGOs.[19] The rise of (voluntary) private standards (or codes of conduct) complementing public and mandatory standards to deal with trade of agricultural commodities, typically monitored through private audits and third-party certification, is perhaps one of the most significant phenomena in the past three decades of expanding agricultural trade (Barrientos et al, 2003; Schuster and Maertens, 2015). Such voluntary private standards can be classified either as own company standards, which affect only the workings and supply chain of a single company[20], or collective standards at both national and international levels, which are available to any number of actors as long as they can fulfil the requirements set by the standard (Henson and Humphrey, 2010). In this review we focus on CS (standards systems) for agricultural commodity production, by which we mean collective (not own-company) standards, subject to third-party certification and auditing processes, where NGOs play an important role.[21] We also include what has been referred to as 'second-party certifications', which are collective standards like third-party certifications but with standard setting and monitoring controlled by an industry/sector, such as large retailers in the case of EurepGAP or the Ethical Trade Initiative in manufacturing (Gereffi et al, 2001; Raynolds and Murray, 2007). These industry-specific schemes also ensure separation of powers between standard setters, certification bodies and accreditation bodies, in what is dubbed a 'tripartite standards regime' (Loconto and Busch, 2010; Aasprong, 2013) .These various private standards are not detached from existing

[18] We distinguish between fair trade, a broad movement for ethical trade, and Fairtrade, the main certification scheme within the fair trade movement.
[19] http://www.intracen.org/itc/market-info-tools/voluntary-standards/standardsmap/
[20] For example, Nestlé AAA standard or Starbucks Coffee and Farmer Equity (C.A.F.E.) Practices and Cocoa Practices.
[21] We use the term certification schemes to refer to programmes that are associated with one or more certifications, with possibly several standards included, each requiring a set of specific requirements (as in the cases of MPS, GlobalGAP or Fairtrade)The CS or standard systems included in this review may or may not use 'labels'. See http://www.isealalliance.org/sites/default/files/private/Researchers'%20Guidance%20Note%20Final%20Apr%202016.pdf

public regulation. In fact usually these standards should or tend to conform to internationally recognised guidelines such as ISO/IEC 17065:2012.[22] A broad definition provided by ISO/IEC states that 'the overall aim of certifying products, processes or services is to give confidence to all interested parties that a product, process or service fulfils specified requirements'.[23] While certification, monitoring and accreditation are fundamental components of CS, their existence also shapes various aspects of supply chains, including the nature of actors and intermediaries (e.g. whether NGOs are actively involved or not), the markets concerned (whether standards are 'mainstreamed' in conventional markets or create alternative channels) and other questions such as traceability and the proportion of the value of a final product that is appropriated by producers (Ouma, 2015; ITC, 2016; Muradian and Pelupessy, 2005). Therefore, as Riisgaard (2009) argues, standards may perform a wide variety of functions depending on actors involved and the nature of agricultural value chains, from cost-cutting to risk mitigating, brand-making, door-opening, awareness-raising among other roles.

In order to understand what CS are and do, therefore, it is important to consider their different origins. This has partly been discussed in the previous section. Most CS for socially sustainable agricultural commodity production have their roots in ideas and movements about ethical trading in Europe and the US, going back at least to the 1980s (Blowfield, 1999; Barratt-Brown, 1993). With supply chains lengthening as a result of the spread of GVC, consumers – and some firms – began to question the pay and working conditions of the workers and producers in LMICs. Ethical (or 'alternative') trade seemed to offer an alternative and by the late 1990s voluntary private standards were firmly established in a number of sectors (Barrientos, 2000; Gereffi et al, 2001). Another source of impetus for certification are food safety and quality standards, aimed primarily at quality assurance, increasingly important for food exports from LMICs to high-income country markets, and especially driven by business interests in the retail sector (Hansen and Trifkovi□, 2014, Raynolds and Murray, 2007; Henson and Jaffee, 2008). These standards may then be complemented by social sustainability standards closer to ethical trade schemes. Thus, as noted in the previous section, since the late 1990s, increasingly powerful retailers have converged to create standards in response to consumers' growing concerns regarding product safety, environmental impact and the health, safety and welfare of workers and animals' (GlobalGAP 2016; Henson and Jaffee, 2008).[24] In this framework there is a trajectory from narrow focus on food safety for consumers at destination to broader consumer concerns about the welfare and health of those producing upstream, i.e. farmers and agricultural workers. The development of schemes that accommodate *multiple* standards is therefore becoming the norm. Web searches of the most widely cited CS show that in most cases these schemes combine various standards usually clustered around three main areas: environmental sustainability, social sustainability, safety and quality.

[22] Which replaced ISO/IEC Guide 65:1996 (http://www.iso.org/iso/home.html)
[23] https://www.iso.org/obp/ui/#iso:std:46568:en
[24] http://www.globalgap.org/uk_en/who-we-are/about-us/history/

While the quantitative significance of standards and CS relative to total agricultural trade is – as noted – difficult to ascertain with any degree of precision, the proliferation of standards and CS is well established. The ITC Standards Map counts 128 sustainability standards for agricultural products alone and it is not uncommon to find several standards per certification scheme.[25] For example, in the standards map Fairtrade appears as having two main standards for agricultural products, namely Fairtrade International Hired Labour and Fairtrade International Small Producer Organisation. A search for GlobalGAP in the agricultural sector at the ITC Standards map yields 4 different standards that are either crop or process-specific. In other words, while CS may be different in some respects, they may also overlap substantially on some of their standards. It is therefore necessary to distinguish between a certification scheme (Fairtrade, MPS, Utz Certified, Rainforest Alliance, etc.) and a standard (more broadly social or environmental standards, and, more specifically, a living wage, the prohibition of certain chemicals, democracy in producer organisations, etc.). The proliferation of standards and CS has not only led to overlaps between them in terms of requirements on producers but also to the phenomenon of multiple certification, whereby producers may be simultaneously certified by various CS with overlapping standards. The implications of multiple audits and principles of evaluation cannot be ignored (Ouma, 2015). There have been cases where a sector develops a new certification that meets criteria of multiple standards in different markets, as in the Chilean fresh fruit sector, precisely to reduce the burden of multiple systems of monitoring and auditing (Aasprong, 2013, p. 93). Within ISEAL, the following members currently operate in the agriculture and food sector:[26]

- Better Cotton Initiative
- Bonsucro (sugar)
- Fairtrade
- Global Coffee Platform (formerly 4C Association)
- IOAS (provides organic accreditation to other certification bodies)
- LEAF (environmental; integrated farm management)
- Roundtable for Sustainable Palm Oil
- SAN (Sustainable Agriculture Network)- to Rainforest Alliance
- Union for Ethical Bio Trade (biodiversity)
- UTZ (coffee, tea, cocoa)

The list above gives a sense of the most important (and most widely known) group within the universe of agricultural standards, but there are others, such as Cotton Made in Africa, Ethical Tea Partnership, Fair Flowers and Plants, MPS, GlobalGAP and several others, which operate in dozens of countries for key agricultural exports. The review did not exclude any of these other schemes and any third- or second-party CS concerned with agricultural products and the wellbeing of producers and workers was included in the search strategy (see Section 3.1.3). The final set of studies reviewed may only refer to a limited set of CS due to availability of eligible impact evaluations.

[25] As above, see http://standardsmap.org
[26] See http://www.isealalliance.org/our-sectors/agriculture-and-food

1.2.2 What certification schemes do and how they do it

Before we unpack what certification schemes do 'on the ground' we need to understand who it is that is being certified. In the case of CS for agricultural and food production, there are three main types of actors that may receive certification:

- Individual farmers (agricultural producers)
- Farmers (producers) organizations
- Export firms/organizations

This review is concerned with those directly involved in the production of agricultural commodities, therefore farmers or producer organizations. The type of interventions may vary according to who is certified. Thus, if a producer organization is the certified actor, the governance of the organization and the distribution of benefits and costs among member producers are critical aspects that can affect the impact of a CS. Likewise, when individual producers are certified, their initial characteristics in terms of scale, technology, resources and knowledge also shape the potential outcomes of certification. CS may also offer different services and forms of support depending on whether a collective group or an individual is certified and on the particular characteristics of the group or individual. Some CS may choose to certify only certain types of producers, i.e. smallholder farmers, as opposed to the entire producer community in any given country.

What then do these CS do? Generally, CS aim to improve upon the effects of 'free' trade by offering better trading conditions, supporting producer organisations to gain better market access, assisting producers to enhance product quality, designing specific interventions or incentives to raise productivity, or a combination of these aspects. Indeed, some CS associated with fair trade and ethical trade movements emerged in response to processes of market liberalization in developing countries (Raynolds, 2000). They also provide markers for product differentiation in increasingly complex and segmented markets where consumers want to know more about the products they consume, where they originate from, how they have been produced and whether they respect the environment and basic human rights (Reardon et al, 2009; Ouma, 2015). In relation to this, CS also act to make labour standards visible either by requirements to meet basic rights (such as a minimum or a living wage) or by assisting firms and workers to improve basic conditions through investments at the workplace. Not all consumers are 'ethical' and some will only be concerned about markers of quality and their own safety. Therefore, CS develop standards and labels to meet all these differentiated demands. It is therefore not surprising that a challenge for any study of certification of agricultural commodities is that standards tend 'to vary in terms of their reach and objectives' and 'there are also major differences regarding the scope of the offering of certified commodities and products' (von Hagen et al, 2010, p. 1).

Overall, in any case, by implementing different bundles of interventions, CS of different kinds are *expected* and often claim to produce positive outcomes across a range of areas of social and environmental sustainability, in addition to quality and the strength and resilience of certified producers in global markets. In cases where social sustainability is the priority, as in the case of Fairtrade for instance,

interventions are thus expected to directly and indirectly empower marginalised agricultural producers, workers and their communities. As Fairtrade summarises it: 'Fairtrade supports farmers and workers in gaining more from trade and through this they are empowered to control their lives'.[27] In cases where social and environmental goals are intertwined, as in Rainforest Alliance, farm productivity and profitability, and the well-being of farmers, workers and their families are seen as intermediate outcomes which, in the long term, and through scaling-up may lead to the ultimate goal of 'creating and maintaining sustainable, resilient rural landscapes'.[28] Some other CS tend to focus on more specific social sustainability outcomes such as production efficiency and profitability and basic labour standards (Bonsucro, 4C-Global Coffee Platform, or Better Cotton Initiative among many other schemes). For Utz meeting the requirements of the standards combined with access to training, and better market connections can result in productivity and quality improvements that enhance opportunities for farmers and workers and may finally contribute to a protected environment (Utz, 2016).CS encompass a wide range of different methods of achieving those goals, although certain types of intervention seem more common across CS (more on this below). An important differentiation has to be made between the act of certifying/licencing itself and direct interventions that precede or follow the certification process. While the development of standards and the act of certification itself may not constitute a conventional development intervention *per se*, the introduction of codified standards, following an auditing and accreditation process, may induce behavioural changes in farmers, notably investments in order to meet requirements, which benefit production conditions and open access to better market opportunities, without any *direct* intervention at farm level by the certifying body. But most CS do require direct interventions at the level of the farm, the producer group or the workers' group. In short, different CS are best understood as bundles of interventions, guided by a variety of theories of change, which will be described in more detail in the following section.

Schemes may also substantially differ in terms of the type of regions and producers they target. For example, Fairtrade tends to focus on LMICs, especially low-income countries in Africa (accounting for 64% of total participants) and Latin America (Faitrade, 2015), whereas GlobalGAP membership is concentrated in Europe, with about 75% of their certified members residing there (Masood, 2014; GlobalGAP, 2016).

CS do not operate in a social, economic or political vacuum. Indeed, they may, and frequently do, coexist alongside additional interventions by NGOs that adhere to the CS social and environmental sustainability standards, as is the case of OXFAM and the Fairtrade certification or TechnoServe (see section on study design for a more elaborate discussion), as well as wider developmental interventions, such as market reforms or the creation and reforms of producer organisations. Therefore, as well as the direct interventions being implemented by CS themselves, there is often some form of external support (by NGOs, donor agencies, buyers) which may have been

[27] http://www.fairtrade.org.uk/en/what-is-fairtrade/what-fairtrade-does
[28] http://www.san.ag/biblioteca/docs/SAN_RA_Impacts_Report.pdf#page=16

leveraged because the producers or groups of producers have obtained a certification. This can affect the interpretation of findings, so the review looked for evidence on these additional factors whether in quantitative evaluations or pieces of qualitative research.

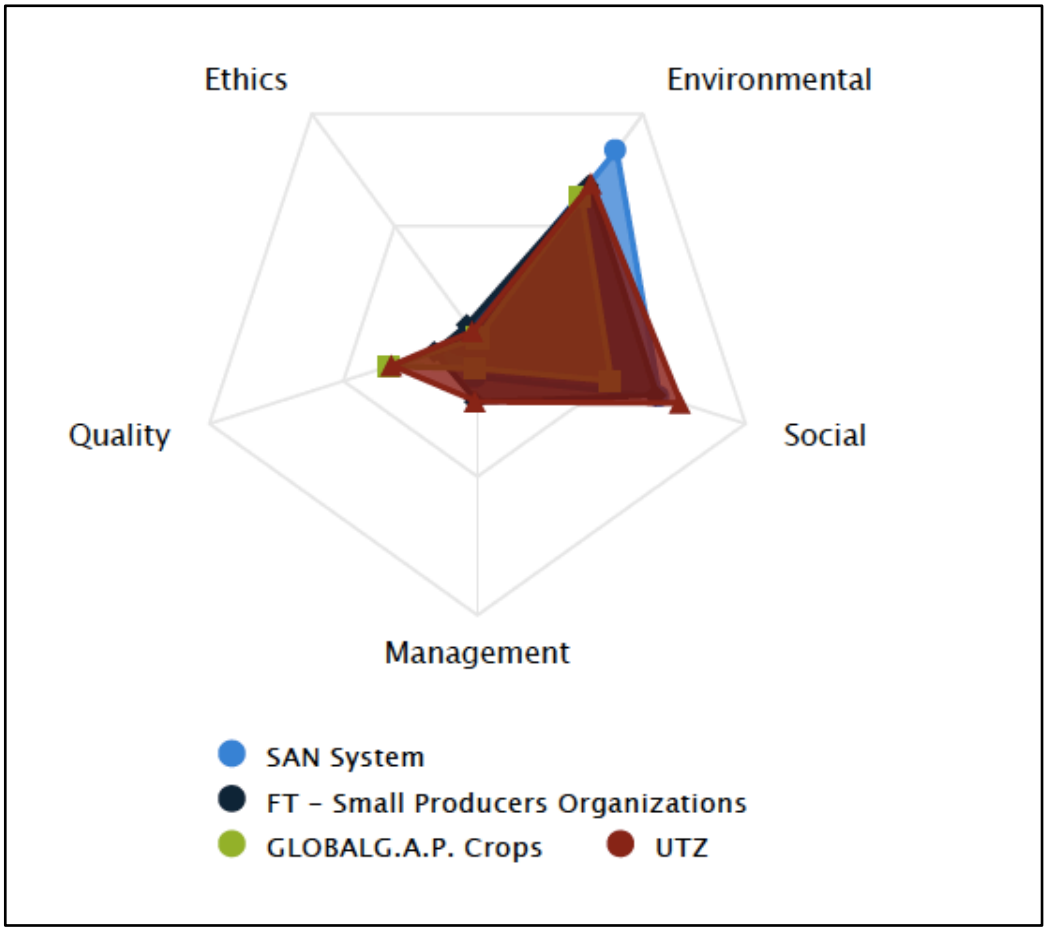
Besides the coexistence of different interventions, some external to the CS, their effects are shaped by a variety of configurations of mediating factors. Among them one stands out: the market conditions and value chain characteristics for each particular commodity that may be subject to a range of standards set and monitored by various CS. CS interventions take place within a given set of market relations and governance structures. Therefore, the nature of the value chain and the dynamics of specific agricultural commodity markets are likely to affect the impact pathways of CS interventions. The linkage between market conditions and effectiveness of CS may also be a function of what type of producers are targeted. In this respect there is also substantial heterogeneity. Apart from the distinction between producers and wage workers as potential beneficiaries it is important to consider the differentiation of producers across CS. Some CS deliberately focus on smallholder farmers, as is the case of Fairtrade with their standards for SPOs (Small Producer Organisations) in coffee production.[29] Other CS have a mix of smallholder farmers in POs and large farmers and agribusiness being certified on the basis of the same standards and supported for production improvements (e.g. Utz, Rainforest Alliance). CS with a focus on quality, safety and demanding social sustainability standards such as MPS-SQ or GlobalG.A.P seem to reach mainly larger-scale producers and agribusiness, partly because the certification costs are very high (Cofre et al, 2012; Bain, 2010).

A further complication for a review of the evidence of effects of CS is that CS increasingly expand their set of standards to qualify for a wider range of markets, products and consumers, and to compete with other CS. This complicates the task of disentangling the specific interventions or standards that really matter in terms of effects on producers' and workers' wellbeing. A scoping survey of CS shows that overlaps can be significant and the wording of standards and codes of conduct are often strikingly similar despite very different histories and modus operandi.[30] For example, most of these schemes apply conventional decent work ILO labour standards as part of their commitment to ethical trade, or share emphasis on 'sustainable farming methods'. Error! Reference source not found. and **3** below illustrate these overlaps through a diamond and a bar chart. Error! Reference source not found. illustrates the overlap between four leading CS: Fairtrade, Utz, SAN-Rainforest Alliance and GlobalGAP Utz, for example, is explicitly active over a range of issues: environmental sustainability, gender equality, productivity improvements; living wage for workers, no child labour and strengthening POs.[31] All four CS focus

[29] A split between Fair Trade USA and Fairtrade in 2012 mainly arose about 'whether the Fair Trade label should only be available to small-scale producers', for coffee markets, with FT USA preferring to include large producers in all commodities (Dragusanu et al 2014, p. 258).
[30] See, for instance for MPS, http://www.my-mps.com/en/certificates-producer, for Fairtradehttp://www.fairtrade.net/our-standards.html, for Utz Certified https://www.utzcertified.org/aboututzcertified
[31] https://www.utz.org/what-we-offer/sector-change/

on standards that are in the environmental and social sustainability areas. **Figure 3** also shows a remarkably similar number of requirements at least for three of the four themes within the social sustainability area of standards, and especially for requirements in relation to employment conditions. The main difference lies in much less emphasis by GlobalGAP on requirements around 'human rights and local communities', compared to the other three CS.

Figure 2*:* **Requirements by standard and sustainability area**



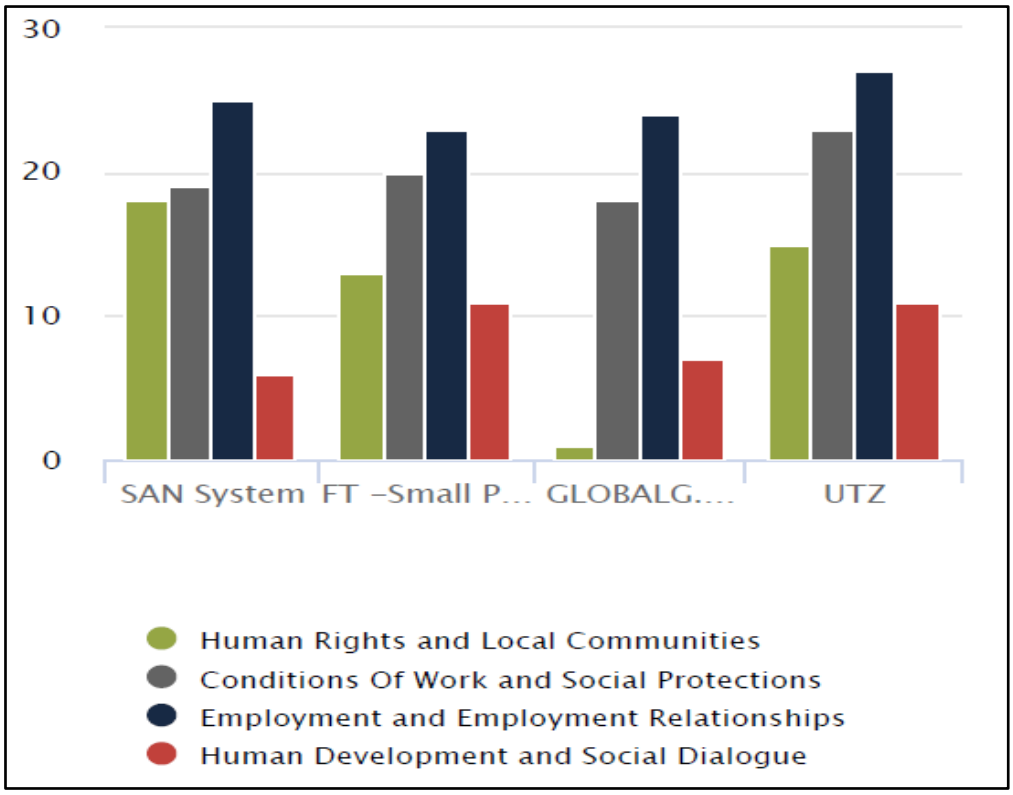
Source: Generated using the standardsmap.org interface.

Overlaps are expected in some cases as different standard systems (e.g. Utz or RA) will have to comply with EurepG.A.P./GlobalGAP standards in order to source European supermarkets. Overlaps also exist in the sense that a particular type of certification can be provided by a variety of certifying bodies/organisations, which may fall under the broad category of voluntary 'social sustainability standards' and conform to broad internationally recognised guidelines such as ISO/IEC 17065:2012, which replaced ISO/IEC Guide 65:1996.32. For instance,Fairtrade certification may be provided by the Fairtrade International (FLO),Fair Trade USA or alternative trade organisations within the WFTO,such as CTM Altromercato. Indeed,the Fair Trade

[32] See http://www.iso.org/iso/home.html for details.

network has evolved significantly in the past three decades and has given rise to a variety of organisations that may share a similar ethos and objectives but may differ in terms of focus, outreach, interventions and auditing processes (Jaffee and Henson,2004; ProForest,2005; Muradian and Pelupessy,2005; Kolk,2005). There can also be various levels of certification by the same certifying body as in the case of MPS[33],depending on what particular standards are applied,with some only focused on environmental outcomes and some including a strong labour standard component (e.g. the MPS-SQ). There is therefore a multiplicity of standards and certifications that often overlap and compete with one another (von Hagen et al,2010).

Figure 3: Number of requirements per theme within social sustainability area: SAN-Rainforest, Fairtrade SPO, GlobalG.A.P. and Utz



Source: Generated using the standardsmap.org interface

A systematic review could in theory be conducted on every single intervention, which could happen under different CS, as in the case of labour standards interventions that are common to most schemes subscribing to ethical trade standards. However, the reality is that most CS operate with bundles of interventions and most studies will report on the fact of being 'certified' and not on single interventions that are part of a scheme. A reasonable judgement can sometimes be made about the dominant intervention, whether it is auditing of labour standards, or training for better farming practices, or the use of price premium. Besides, seemingly similar interventions may

[33] The name MPS comes from 'Milieu Project Sierteelt', but is generally no longer used an acronym.

be structured and implemented differently in different places and at different times, encompassing different intervention components. For example, technical assistance and capacity building for better farming practices or to improve organisational performance may be implemented with a variety of intervention components, such as direct extension services or simply through training toolkits distributed to POs. This makes the analysis of the causal chain particularly complicated because endpoint outcomes may be attributable to a bundle of interventions without sufficient evidence on which particular intervention component is more effective. For example, in the case of MPS-SQ, is the certification more effective because of the enforcement of labour standards or because of the quality standards generally imposed through MPS-type standards and their spill-over effects on other intermediate outcomes? Most impact evaluations will find it difficult if not impossible to disentangle the specific effects of these different interventions under the same scheme. At the same time, most CS will consider that what matters is the specific mix of interventions,for instance including various forms of capacity building for producers, and not any one intervention in particular. Depending on the level of detail reported by available studies some insights into the key causal mechanisms either through quantitative or qualitative evidence would be considered to assess Review Question 2 (see the section on objectives for questions addressed in this review).

Finally,the description of the interventions under CS must include the fact that the process of certification involves costs. This is indeed a contentious issue when producers are asked to evaluate the benefits of certification. While there is generally a perception of potential benefits, certified producers tend to perceive such benefits in relation to the cost of obtaining them. In most CS,there are two main types of costs:

- implementation costs so that standards can be reached, which will be higher if the standards required are very demanding in comparison with conventional production systems; and
- direct certification costs, resulting from the process of auditing and granting of certification by a third party or certifying body.

In addition to these two main types of costs,it is also worth considering opportunity costs arising from the diversion of resources to certified production and away from other activities. The incidence of all these costs is of vital importance to understand their impact. According to information compiled by the ITC on over 100 different certification systems, producers alone bore the full implementation costs in over 60% of certifications and the full certification costs in over 50%. In around a quarter of certification systems both types of costs were shared between direct producers and other supply chain players (ITC,2016). Moreover, the level of costs can differ substantially, which can help determine the possible incidence. For instance, GlobalGAP certification can be extremely expensive and the costs is always borne by a larger supply chain player, such as an exporting company, or by a large producer organisation (Bain,2010). Some CS like the Roundtable for Sustainable Palm Oil make exceptions and cover 100% of certification costs for smallholders (but

not for other larger growers or mills),as part of their commitment to promote social sustainability in palm oil production.[34]

1.3 How the interventions might work – a theory of change

A major challenge for this review was the way in which different CS that aim to improve the welfare of agricultural producers and workers in agriculture differ in their model of intervention and in their theory of change (ToC). For example, Fairtrade focuses on fair prices and market access,MPS is mainly about sustainable quality and social standards, and UTZ Certified, while similar to Fairtrade schemes in the sense of its broad aims, works in terms of improvements in farming practices, productivity and quality rather than price mechanisms**.** Moreover,each certification scheme may also incorporate different grades of certification, as in the case of MPS for flowers, GlobalGAP and all of its different standards, or the different standards applied by Fairtrade to SPOs (Small Producer Organisations) or HLOs (Hired Labour Organisations, i.e. large-scale plantations). As explained in the previous section, although there is convergence among CS in terms of the sustainability areas they aim to cover, there is a whole range of standards and requirements that are bundled within each programme.

Furthermore, each ToC may contain a substantial number of aspects and causal chains, considering a range of contextual factors. As the ISEAL researcher guidance notes, 'Standards' Theories of Change are complex because standards operate in complex environments where lots of different actors also operate. They are also complex because they address complex issues such as workers' rights, child labour, deforestation, sustainable livelihoods and so on' (ISEAL,2016, p. 6). Some contextual factors are shared by different CS but the degree of complexity across ToC may also vary. Given the wide variety of CS, their intended outcomes and methods of intervention there is no *single* theory of change that is valid for *all* types of CS. There have been attempts by researchers to develop a ToC valid for more than one CS (Nelson and Martin, 2011 and 2013). Indeed in 2009, as reported by these authors, 'sustainability standards had yet to articulate their own theories of change […] although this situation has now changed as a result of the ISEAL Impacts Code and with contributions from this research project', including subsequent studies that drew on reports by the Natural Resources Institute (NRI) to develop ToC for impact assessments. Some CS have also recently produced an explicit theory of change, and readers may benefit from consulting the ToC developed, for example, by Utz Certified, Rainforest Alliance and Fairtrade as indicative examples.[35] In recent years, most ISEAL members have already produced

[34] See http://www.rspo.org/about/how-we-work

[35] See, for example, Fairtrade ToCs, published in 2013, Utz ToC diagram from its website, and Rainforest Alliance ToC diagram, respectively at:
http://www.fairtrade.net/fileadmin/user_upload/content/2009/resources/140112_Theory_of_Change_and_Indicators_Public.pdf
https://www.utz.org/?attachment_id=4236
http://www.san.ag/biblioteca/docs/SAN_RA_Impacts_Report.pdf#page=16
ISEAL (2016, p. 6) also provides a list of CS with links to their ToC or similar attempts at linking inputs and impact

a ToC or at least a set of hypothesised causal links between inputs and impacts (see links in ISEAL 2016). As part of this process of mainstreaming the impact code, ISEAL have also produced a ToC and routinely publish drafts and discussions of ToC for individual member organisations.[36] GlobalGAP, on the other hand, does not appear to have developed an explicit theory of change, although its focus on food safety is frequently stressed and its emphasis on producers' and workers' safety and wellbeing also strong.

Drawing from these various examples of ToC, the claims made by some CS and the key interventions they propose, we produced for the protocol a simplified synthetic ToC that summarised the key linkages in the causal chain between broad *types* or groups of interventions, intermediate outcomes and endpoint outcomes, bearing in mind that the focus of this review is on *socio-economic outcomes*, and in line with some earlier attempts such as Nelson and Martin (2011). Some of the organisational ToC mentioned above, particularly the one developed in 2013 by Fairtrade, may be more complex and multifaceted than the synthetic ToC we propose here. This is because CS like Fairtrade also focus on actions and advocacy among consumers to expand the market for Fairtrade certified products and generally the values of Fair Trade. They also include environmental standards as part of the broad canvass of sustainable outcomes, like most other CS found in this sector. The focus of this review is, however, on the role of standards and interventions that more directly affect the wellbeing of producers and workers involved in the production of certified commodities. The aim is not to evaluate the work of all these different CS on all the outcomes they monitor, rather to evaluate and synthesise the existing evidence on *socio-economic outcomes* associated with interventions under CS as defined in this review. This was necessary to keep the review manageable and allow a consistent framework that can be applied to a wider range of CS. Indeed, in the course of the review, it was soon clear that the scope was broad enough, given the range of CS and types of interventions and outcomes, making the screening process an arduous exercise.

Below are illustrations of how different types/groups of interventions, used by different CS, may affect intended outcomes, and therefore the assumed causal chains. Here we highlight the direct interventions in addition to the usual standard setting process, which creates requirements for producers to meet in order to be certified. This, in itself, is the key intervention that distinguishes CS from other sources of support and intervention affecting agricultural producers and workers. There are of course many possible ways of classifying interventions depending on what level of analysis one chooses or which CS are analysed. Given that this review covers a significant range of CS, it is difficult to agree on a straightforward taxonomy of interventions. For example the International Standards Map developed by the ITC distinguishes different forms of 'support' broadly understood as interventions developed by CS to facilitate and manage the certification of agricultural products: documents and guidance tools, equipment, technical assistance for certification,

[36] See http://www.isealalliance.org/tag/theory-of-change

technical assistance beyond standards' requirements, access to finance, price and premium (ITC, 2016).

One option is to highlight the most common interventions considering different types of CS, from more quality and market oriented to those more oriented to ethical standards and sustainability. **Figures 2** and **3** above illustrate different sustainability 'areas' and within each a number of themes that characterise groups of interventions, as reported by the International Standards Map (ITC, 2016). Section 4.1.2 reports on six typical types of interventions frequently reported in evaluations of different CS: price interventions (minimum/guarantee and additional premium); premium as additional payments to support producers and their organisations; other market interventions to reduce volatility and/or improve market access to more remunerative/more stable markets (output markets) or to input markets (seed, credit); support for better organisational and production management; support for better agricultural practices for quality, productivity and health and safety in production, through technical assistance and guidance tools; labour standards through auditing and monitoring.

However, it is possible conceptually to organise these six types of interventions in broader groups with clearer conceptual boundaries. This more general classification also helps synthesise different possible theories of change underpinning different types of CS. Therefore, the main interventions that come with the certification process may be grouped around four main blocks of related interventions:

1. **Capacity building interventions.** These include the following:
    a. Capacity building for value-chain upgrading, such as:
        i. Training and extension service for better farming practices to improve quality, productivity and/or food safety.
        ii. Assistance for professional farm management, through training visits or materials.
        iii. Assistance/training for professional PO management, especially in relation to coordination between producer member, exporters and buyers, quality/grading checks and other tasks that can increase the value of traded products.
    b. Support to POs and business to improve the governance of their organisations and democratise decisions over use of premium. This form of capacity building is expected to empower the members of these POs and make them more sustainable and accountable.

    The first type of capacity building, which is also what we refer to as 'good agricultural practices' in the taxonomy used in Section 4.1.2 is very common across a wide range of CS, whereas the second type is typical of Fairtrade.

    Altogether these forms of capacity building are expected to:
    1. On the upgrading dimension:
        - ✓ Improve farm management through behaviour/attitude changes combined with more resources

- ✓ Lead to investments in improvements in quality and/or productivity of traded commodities,
- ✓ thereby commanding better prices, and
- ✓ thus contributing to higher farm incomes and overall household income, which may also positively affect wealth and household investments in education and health.

2. On the governance dimension:
   - ✓ Improve professional and democratic management of organisations, which can strengthen organisations (POs or agribusiness) in terms of their legitimacy, participation and capacity to negotiate,
   - ✓ which can lead to members' empowerment and access to better services and more remunerative markets thanks to better reputation and organisation.

2. **Market interventions**. These include two main groups:
   a. Price interventions (output markets), such as:
      i. Price floor/guarantee to cover basic production costs and protect producers from market fluctuations and slumps.
      ii. Price premium accruing to producers if certain attributes are achieved.
   b. Other market interventions in input and output markets, namely:
      i. Some form of pre-finance or credit.
      ii. Longer-term or more stable contracts.
      iii. Access to alternative and/or additional possibly more lucrative markets if certification requirements are met.

   They are expected to:
   - ✓ Contribute to higher and more stable producer prices, which in turn can result in higher net profits for agricultural producers, assuming they are not offset by high certification costs.
   - ✓ Protection from price volatility can improve reliability of supplies and/or predictability of sales, resulting in greater income stability, profitability and reduction in risk and vulnerability to shocks. These effects can result in higher incomes and consumption at household level as well as reduced vulnerability to poverty at the end of the causal chain.
   - ✓ Support in input markets can also improve producers' capacity to invest and improve production conditions and productivity.
   - ✓ All of the above contribute to strengthened market power and negotiation capacities of producer organisations and ultimately contribute to their members' empowerment.
   - ✓ All of the above may also result in asset building, both productive assets in farming and also household assets, which may also positively affect wealth and household investments in education and health, as well as improve resilience against shocks.

3. **Premium-funded social, community and business investments.** This is a special type of intervention that straddles the boundaries between market interventions and capacity building and is particularly important in one of the

leading CS, Fairtrade. It is singled out as it can generate its own causal chain for key socio-economic outcomes of interest in this review. This premium is also sometimes referred to as a 'social premium' or 'community premium' because the price premium offered on top of the market price to a PO or a plantation can be invested in a variety of assets/infrastructure leading to possible positive outcomes, such as:

- ✓ Better education and health access and/or other outcomes, which may also positively affect wealth and household investments in education and health.
- ✓ Higher incomes if economic infrastructure/assets improve production and marketing conditions.
- ✓ Empowerment via strengthened beneficiary organisations.
- ✓ Better working conditions, when premium-funded investments directly affect the non-wage conditions faced by agricultural workers.

4. **Labour standards**. Their implementation, through monitoring and auditing of working conditions in farms, can directly impact workers' wellbeing through ensuring payment of living/better wages, and better working conditions, especially when health and safety conditions improve and affect workers' health.

From this list above, and based on exploratory analysis of the best known CS, the most common forms of interventions within the wide range of standards consulted are the following:

- **Development and auditing of standards:** this is by definition what characterises CS intervention; the requirements for each agreed standard usually require investments on the part of producers to improve production and quality to meet agreed standards. This includes the important category of labour standards and the auditing process that accompany them. All CS do this in one way or another.
- **Assistance for good agricultural practices**, to meet agreed standards and empower producers to produce more, better and more safely.
- **Assistance to POs to improve performance** and contribute to value chain upgrading as well as to higher quality governance and accountability.
- **Price interventions**, whether floor price (Fairtrade) or some kind of premium (Fairtrade, via price premium, and others such as Utz).

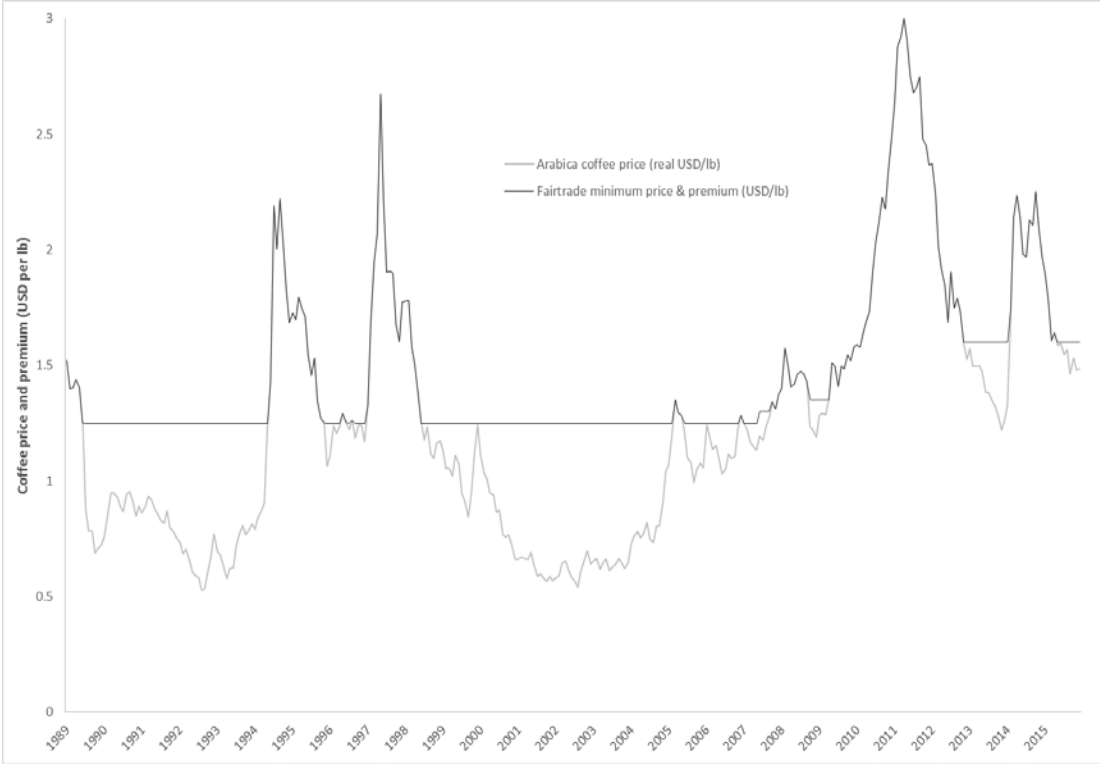
All CS included in this review have a system of *monitoring and auditing standards*. Once standards are agreed the pathways to impact on welfare of producers and workers can take four main forms:

- If standards grant access to more remunerative markets or guaranteed buyers, farm income can become higher and more stable. This could potentially, or not, trickle down to conditions for wage workers employed by certified farms.
- If standards include good agricultural practices designed to improve conditions in the production process and thereby the health of workers and producers, the endpoint effect could be improved health and reduced vulnerability to sickness, for example.

- If standards specifically refer to working conditions, then wage workers in certified farms directly benefit from how compliance affects their wages (e.g. a living wage) or non-wage conditions.
- If standards entail compliance with practices that improve productivity and/or quality, producers can benefit from higher yields and higher quality thereby commanding higher prices in conventional markets or even access to niche markets where quality is highly valued.

So, in this latter pathway to impact a standard that is audited and leads to compliance entails investments by producers (with their associated costs), reflecting behavioural changes, that may improve production and working conditions and improve both environmental and social sustainability.

Figure 4: Comparison of Fairtrade minimum price and Market prices for Coffee: 1989-2015



Source: Own elaboration from a combination of Fairtrade data (https://www.fairtrade.net/standards/price-and-premium-info.html) and World Bank 'pink sheets' commodity price data. A similar graph appears in Fairtrade Foundation (2012).

The effectiveness of *price interventions* depends on several factors but in particular, for example, whether (a) the floor price is above ongoing market prices; (b) price premia is sufficiently large enough to trickle down to producers, particularly in the case of smallholders since a small percentage of premium added to sales may have a very marginal impact on total household income, especially if only part of the production is sold as certified produce. For the Fairtrade floor price, the data since

the 1990s show extended periods in which the intervention did *not* really apply as international prices exceeded the floor price (**Figure 4**). This floor price has been revised over time based on estimates of reasonable production costs. Therefore, in years when the floor price is irrelevant the impact of price interventions may potentially be marginal (Dragusanu et al 2014). When the intervention entails a premium, the question is whether the premium is directly passed onto individual producers or goes to producer organisations, which then decide on the uses of a collective premium, usually in the form of investments in social and productive infrastructure to benefit the organisation or the communities where producers are located.

The various forms of capacity building and direct assistance to individual producers or POs usually complement the system of auditing of standards and are designed to help producers meet such standards and achieve upgrading in the value chain. The interventions may be directed to individual producers or to groups of producers and may be addressed to producer organisations when these are the main focal points of the interventions, as in the case of most CS working with smallholder farmers. The pathway to impact from assistance depends on the *type of assistance*, whether this is designed to improve the agricultural practices to increase yields, or to reduce environmental damage, to reduce vulnerability to chemicals and disease, or even to improve the organisational capacities of producer organisations, including their internal governance mechanisms,[37] as well as on whether such assistance is continuous or granted as a one-off to help producers attain the standards necessary for certification.

Figure 5 presents a simplified synthetic theory of change, which captures the overall logic of interventions under CS, according to the four broad categories listed above. As discussed, this is synthesised from multiple theories of change from some of the most prominent CS types. The synthetic theory of change was developed to be broad enough to be able to capture all intervention methods we are going to encounter under various CS. It summarises causal chains and key assumptions for four different broad intervention types, namely interventions around farm practices, around prices, markets and purchasing agreements, around labour standards and around the use of the price premium. These distinctions respond to the logic of potential causal linkages between these interventions and some of the key socio-economic outcomes, which this review focuses on. This synthetic theory of change illustrates the difficulties inherent in aggregating results on effectiveness over a heterogeneous body of schemes and interventions.

A key aspect of any theory of change is a listing of the assumptions that must hold at each step along the causal chain for interventions to have their desired effect. If

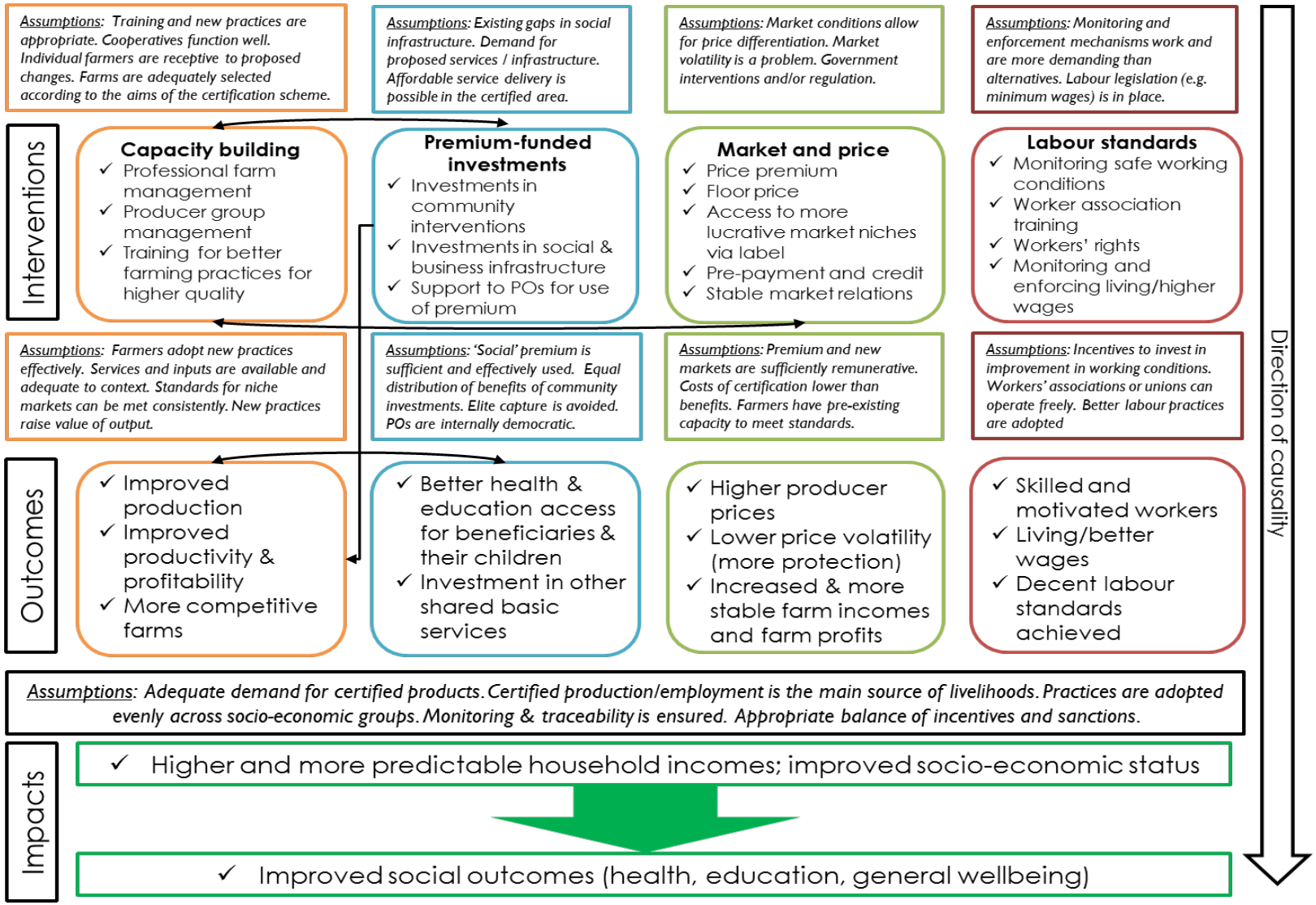
[37] Assistance is provided to POs so that they can organise themselves to respect the principles of democratic governance, The Fairtrade website states under the standard for SPOs the principle of democracy, i.e. 'Democracy. Profits should be equally distributed among the producers. All members have a voice and vote in the decision-making process of the organization.' http://www.fairtrade.net/standards/our-standards/small-producer-standards.html

assumptions do not hold effects may be diminished, skewed, or entirely absent. In the worst case there may even be unintended adverse effects on producers or workers. However, assumptions also differ in their importance for different interventions and thereby certification types. For instance in some cases farmers' pre-existing capacities and therefore self-selection into the scheme (as in quality-oriented schemes) are more important than others, particularly where the costs of certification or compliance are very high or if CS organisations are selective about the kind of POs they want to work with. In other cases assumptions about the distribution of benefits among members of a group matter more when beneficiaries are targeted in groups (as with Fairtrade premium for small producer organisations) than when they are targeted individually. The distribution of benefits may also not be equal between workers and employees, where large employers are targeted, or there may be differences between different types of workers.

Overall most CS that have produced a ToC (Fairtrade being notable in the detail they offer), acknowledge and stress the importance of contextual factors to explain impacts, i.e. improved incomes, environmental sustainability, fairness in business practices and trade, enhanced gender equality and dignity and voice for small producers and workers. The influence of contextual factors increases as we move from outputs to impacts and the CS influence decreases along the chain.[38]

[38] See http://www.isealalliance.org/tag/theory-of-change

Figure 5: Simplified synthetic theory of change



1.4 Why the review is important

This systematic review addresses the extent to which, and under what conditions, interventions under various CS for agricultural commodity production result in higher socio-economic welfare for agricultural producers and workers in low- and middle-income countries (LMICs) – questions about which there is an ongoing and as yet unsettled debate.

Section 1.1 above has discussed some of the existing evidence about the overall impact of interventions under CS, which presents a wide range of reported results. There are also useful reviews that have mapped the various codes of conduct, especially for wage workers, and the way these incorporate issues of gender and how they operate, but these tend to be focused on the nature, process and actors in these schemes rather than on their impact (see Barrientos et al, 2003 for a seminal study of this kind of mapping). Overall, there seems to be a consensus among those who have studied CS for some time that the evidence is mixed at best.

There have been some attempts to review the evidence more systematically. A study by the International Trade Centre (2011), one of a four-part review series on CS, for instance seeks to present the overall findings of the relevant literature using systematic review methods. However, the study uses vote counting, rather than a meta-analytic method that takes effect sizes into account, to synthesise the evidence and no information on effect sizes is presented. The study also provides no evidence on Risk of Bias (RoB) for individual studies included in the review. The search methods used by the study also cast doubts on how comprehensive its literature coverage is. Searching seemed to have been limited almost exclusively to two databases containing only academic journals.

Similarly, Blackman & Rivera (2010) also use systematic review methods to synthesise the available evidence on sustainability standards. However, this review suffers from similar issues as the study by the International Trade Centre, namely the reliance on a simple vote counting method, a lack of detail on quality appraisal and an unconvincing search strategy. In short, there is need for a high-quality systematic review using more sophisticated methods of searching and synthesis under established protocols like those set by Campbell Collaboration.

The abundance of literature on Fairtrade has also led to some specific reviews focused on this CS. For example, a literature review was commissioned by the Fairtrade Foundation to map and analyse the impact of Fairtrade certification (Nelson and Pound, 2009), while a similar compilation was conducted by Vagneron and Roquigny (2011). Further, Terstappen et al (2013) undertook a systematic scoping review on the social dimensions of Fairtrade, focusing on gender, health, labour and equity in particular. Overall, the three reviews present an account of the existing research, identify some methodological issues (Terstappen et al, 2013; Nelson and Pound, 2009), and make future research recommendations (Terstappen et al, 2013; Vagneron and Roquigny, 2011). None of these reviews, however, provide sufficient information on the searching and synthesis process, nor do they systematically assess the quality of the studies they include. Moreover, they do not conduct a

statistical meta-analysis of effect sizes or a rigorous and exhaustive synthesis of the qualitative evidence.

Another reason for conducting a systematic review with this broad scope, apart from expanding our knowledge of evidence beyond well-known cases such as Fairtrade, is the general preliminary assessment that the main bulk of studies is still characterised by evaluation designs vulnerable to validity threats, while the description of data collection and analysis tends to be poor, preventing assessments of the quality of the evidence (Terstappen et al, 2013; Cramer et al, 2014, and Ruben, 2013). Some of these critical appraisals also point to an existing bias towards giving more attention to independent agricultural producers as opposed to wage workers (International Trade Centre, 2011, p. 19). Therefore, the need for a systematic review with an inclusive framework, which identifies this expanding body of literature and critically appraises its quality, is clear and timely.

Who are the potential stakeholders in this review? The results of this review should be immediately relevant to both policy and practice, since they can provide guidance and lessons to certifying organisations, such as those who are members of the ISEAL Alliance, sectoral/industry codes of conduct (such as MPS or GlobalGAP) and broadly ethical trading partners. The main lessons can be in relation to effectiveness and the kinds of barriers and facilitators and contextual factors that matter most. CS may be interested in knowing more about the relative merit of different components of their interventions, so that selectivity and sequencing may be improved. Key lessons may be learned on methodological issues, especially the standards expected for impact evaluations as well as the required standards for good reporting of evidence, whether quantitative or qualitative. In this respect, it is hoped that the results of this review will contribute to on-going academic debates around the effectiveness of agricultural CS and can help guide future research into areas where the evidence is either weak or ambiguous. CS can engage in a productive debate about what evaluation approaches are both rigorous and feasible and especially how to improve the reporting of research results so that future systematic reviews can produce useful syntheses of the available evidence.

Certifications are also becoming increasingly important to successful entry into GVCs, and are therefore receiving more and more attention in development policy circles. In addition, some of these CS, for example Fair Trade schemes (e.g. Fairtrade, Fair Trade USA as the main ones), also receive public funding from government agencies aiming to improve rural livelihoods (e.g. DFID) and organisations that provide financial or technical support to such certification efforts can also benefit from this comprehensive effectiveness review. The results will of course also be of direct interest to corporations engaged in buying agricultural produce from LMICs, and can contribute to debates around corporate social responsibility (Mezzadri, 2014). Since many CS are centred on providing key information on how products reach the market and how they have been produced, consumers may indeed be important users of this review, especially through its outreach via conventional and social media. In particular, consumer groups or associations may also be interested, as they can gain knowledge to better inform

their campaigns and priorities. Indeed, growing scepticism among consumers due to proliferation of standards and labels calls for a review of the evidence. If evidence of impact is not convincing, agricultural producers may become increasingly resistant to adopting certification. Certainly we hope the review to be of use for agricultural producer organisations and workers' organisations, which invest resources in the certification processes of their members, as well as for individual agricultural producers who also invest in certification to achieve positive outcomes.

2. Objectives of the review

The primary objective of the review is to evaluate and – wherever possible – synthesise evidence on the effects of certification schemes for sustainable and safe agricultural commodity production on key *socio-economic outcomes* at the level of the individual producer and/or worker, and the key barriers and facilitators mediating measured effects.

Therefore, the review set out to answer the following questions:
Primary Review Question (Review Question 1):

1. What are the effects of certification schemes for sustainable agricultural production, and their associated interventions, in terms of endpoint socio-economic outcomes for household/individual wellbeing, such as income (including farm income), consumption, assets, working conditions, education, health (including nutrition and food security) and empowerment in low and middle income countries?

Subsidiary Review Question (Review Question 2):

2. Under what circumstances and why do certification schemes for agricultural commodities have the *intended* and/or *unintended* effects? What are the barriers and facilitators to such certification's *intended* and/or *unintended* effects?

Wherever the data allow to, this systematic review reports on *both* intermediate and endpoint outcomes, since many CS are primarily focused on and interested in these intermediate outcomes, which may often be only *one* of many contributors to the ultimate or endpoint outcomes (Ton et al, 2014). Indeed, most ToC of CS that have developed them clearly state that the sphere of influence of their intervention concentrates around intermediate outcomes and that contextual factors are many and too complex to establish more direct causal mechanisms with long-term impact outcomes. This review focuses on objective dimensions of wellbeing given the intrinsic difficulty in analysing subjective dimensions across diverse contexts with a variety of interventions. There is no one single measure of 'objective' wellbeing, thus Section 3.1.4 proposes a number of indicators that are normally associated with improved socio-economic status, even if not always fully consistent with subjective measures. The simplified integrated synthesis developed in this protocol takes this into account by considering different theories of change embedded in different certification schemes and the limitations of available methods in establishing clear causal attribution on outcome effects to particular certification schemes and their interventions.

The subsidiary review question is important for a number of reasons. First, as stated above, this review synthesises and evaluates evidence of what works where, for whom and under what circumstances. Second, there is an abundance of qualitative and mixed-method research in impact evaluations of CS, which can provide valuable evidence for the subsidiary review question, even if not enough to address the primary review question. Indeed, as the results section will show, the literature on CS is dominated by this kind of literature. In fact, the vast majority of published academic work on CS is largely qualitative in nature. Third, the subsidiary review question can help explore the relative roles played by the type of CS and their interventions and other specific contextual factors, or whether local/national context dimensions are more or less important than aspects of the supply chain beyond national boundaries. Fourth, while the ToC of most CS are explicit about the expected positive outcomes, there seems to be a gap in understanding unintended outcomes, whether negative or positive, and the circumstances in which these arise. Fifth, there are distributional dynamics and contextual aspects that can shape the effects of CS, with possibly uneven distribution of benefits and costs among participants. To the extent that studies report on these matters, the synthesis for RQ2 will also explore these questions.

3. Methods

3.1 Criteria for considering studies for this review

3.1.1 Types of studies

This chapter on methods is based on the protocol developed for the review and published in July 2015 (Oya et al, 2015). As per the protocol we included studies published no earlier than 1990 and available in either English, German, French, Spanish or Portuguese. We included both peer-reviewed and non-peer-reviewed material and no study was excluded based on its publication status. The methodological approach proposed in the protocol was developed in light of the following main considerations.

First, there are inherent challenges in a systematic review that includes a range of different interventions, usually bundled within the same programme (or CS) and a variety of intermediate and endpoint outcomes. This review faced the challenge of trying to synthesize effects for a wide range of intermediate and endpoint outcomes (e.g. incomes, prices, schooling, etc.) for a range of certifications (Fairtrade, GlobalGAP, etc.), each characterised by a specific bundle of interventions (access to certified markets, price premiums, training, credit finance, etc.). Therefore it was expected that the breadth of the review and the heterogeneity of possible documents would produce a wide range of study designs and types of publications to be considered for inclusion.

Second, the literature on private standards or VSS and CS is vast, as became clear in the searching process (see below). As well as empirical research into impact, the literature identified in the search process includes a substantial amount of literature reviews, attempts to analytically frame the rationale for the use of private standards, and syntheses of different types of studies on different CS over a range of

agricultural commodities. Whereas these latter studies were generally outside the scope of the review, some of the discussions therein were useful to understand the history and logic of CS in agricultural trade and to inform a synthetic theory of change, based on the different types of interventions and pathways to impact.

Third, given the nature of RQ1 and RQ2 and the expectation that many of the contributions to literature around the impact of CS use qualitative research, the review adopts a theory-based, mixed methods approach and includes a broad range of evidence from both quantitative and qualitative research. Although only quantitative studies are used to estimate impact (i.e. RQ1), it is clear that there are multiple contextual factors underpinning causal pathways, hence the importance of consulting a wider range of sources to inform RQ2.

The types of studies we considered includable differ for RQ1 and RQ2. The respective methodological criteria are presented below.

Review Question (1)

For RQ1 we only included studies whose design allows for an estimation of the impact of CS on the outcome measures of interest (see below), as well as for a causal attribution of that change in dependent variables to participation in a certification scheme. The effectiveness review therefore focused on rigorous impact evaluation studies using experimental and quasi-experimental designs, including controlled before and after (CBA) studies with contemporaneous data collection and with two or more control and intervention sites, as well as *ex post* observational studies with non-treated comparison groups and adequate control for confounding.[39] Adequate controls for selection bias and confounding factors are vital if we are to have confident in the claims studies make about the causal effects of interventions. Studies following a pipeline approach can therefore only be included if they employ additional methods of ensuring group equivalence, such as matching on observables. To answer Review Question 1 the review included studies with a comparison between producers or wage workers receiving a relevant intervention with a control group that receives no intervention. Studies could also compare several different CS at once, and there may not be an untreated ('pure') control group, but such studies are then not included in the meta-analysis.

As discussed below, we did not find any relevant experimental studies, thus obviating the need to discuss inclusion criteria for such studies. We are aware of two different RCTs looking at the impact of certification which are underway, but findings were not yet available at the time of writing (for baseline reports, see Kumar et al, 2015, and Neilson & Toth, 2016). The remainder of this section thus focuses on

[39] While the use of evidence from quasi-experimental studies in meta-analysis has been criticised, the lack of experimental evidence in many areas of international development research, including in research on the impact of CS, has meant that there are few practical alternatives to the careful aggregation of such evidence, if we are to proceed with quantitative research synthesis (see also Duvendack et al. 2012). Moreover, there is also a debate as to whether randomisation is *required* to solve selection problems, and an imperfectly implemented RCT may not be methodologically preferable to a well-designed quasi-experimental study (Deaton and Cartwright 2016).

quasi-experimental designs. To be considered includable, a quasi-experimental study design had to feature (at least one) control group and use analytical methods that ensure adequate controls for selection bias and confounding factors. We therefore included both controlled before-and-after designs, which collect data at two or more different points in time, as well as *ex post* controlled studies with contemporaneous data collection across treatment and control groups.[40] In these comparisons, individuals were considered as associated with interventions under CS where there are groups of certified agricultural producers or workers, producers' organisations or trade unions, or when residing in geographic areas with a very strong presence of certifying organisations.

In practice, studies using data collection at two different times do not necessarily collect data at baseline and endline, as is common in the evaluation of time-bound projects. Rather treatment groups are frequently already certified, or in the process of obtaining certification, during 'baseline' data collection. Includable analytical methods for quasi-experimental study designs are then matching methods (either on covariates or, more commonly, propensity scores), difference-in-difference measures (which may be combined with matching), instrumental variables, and a variety of other multivariate regression techniques, including treatment models, that provide adequate controls. We would have included regression discontinuity designs, though no relevant studies of this type where identified during the search process, perhaps owing to the potential difficulty of implementing them in evaluating certification schemes into which participants self-select.

Review Question (2)

Unlike RQ1, RQ2 does not seek to establish the impact of CS, but to illuminate the barriers and facilitators mediating the effects of CS. To do so, three key themes of interest were identified:

- implementation dynamics (or how interventions actually worked and with what constraints);
- distributional dynamics (or to what extent some groups may have benefited more and why); and
- a range of other barriers and facilitators, particularly external contextual factors that shape the strength of effects and their distribution and sustainability.

To address RQ2 we included purely qualitative studies on CS, as well as qualitative research which was either part of quantitative impact evaluations or mixed-methods evaluations. In all cases includable studies needed to meet the following methodological criteria:

1. The research question or objective had to be clearly reported.
2. Data collection methods and, where appropriate, sampling procedures had to be clearly reported.

[40] We agree with Ravallion (2007) that the terminology of 'treatment' groups is unfortunate in the social sciences, but unavoidable due to its ubiquity in the literature.

3. The study had to provide evidence based on primary data collected from CS beneficiaries, facilitators, implementers, extension agents, auditors or experts analysed using qualitative methods.
4. The study had to provide substantive evidence on at least one of the key themes of interest (implementation dynamics, distributional dynamics and contextual factors).

By substantial primary evidence we mean that the material needed to provide at least one of the below:

1. A thick description: This means a detailed description of the relevant context together with an analysis of how this context affects at least one of three key themes.
2. Entire sections devoted to analysis of at least one of the key themes. Studies providing only mentions, or non-analytic descriptions of general or historical context were not included.

As a result, studies which did not incorporate reliable and substantial factual (primary) evidence were excluded from the review. This applied to a large number of studies related to advocacy research, in the form of briefs, books, reports and articles, which, this review found, are abundant in the CS literature. On the other hand, impact evaluations and other research with substantial primary evidence on the relevant themes were included. ~~.~~Finally, it is worth highlighting that although 'thick descriptions' (to somewhat oversimply Geertz, 1973), ethnographic work and triangulated analyses were the primary targets for this kind of evidence, non-ethnographic work that lacked triangulated analysis was also included, as long as to satisfied the four above mention methodological criteria. While we undertook a detailed quality assessment of studies included for RQ2, assessing study quality characteristics such as data triangulation (see Section 3.3.3.), this assessment was not used to further exclude studies, but to provide an overall picture of the quality of the studies included.

3.1.2 Types of participants

The review considered two broad categories of participants:

- *agricultural producers (farmers)* and
- *wage workers*

Where 'wage workers' refers to anyone doing agricultural work in return for pay, regardless of whether the arrangement is based on a formal contract, and irrespective the business scale of the farm, i.e. this category includes the employees of large-scale plantations and workers hired by smallholder farmers, producer organisations, or processing units. Differences in employment status were taken into consideration during the data synthesis. Both types of beneficiaries should live in low- and middle-income countries (LMICs), as defined by the World Bank, at the time the intervention was carried out. The way the target group is defined depends on the intervention and the type of study. Some CS operate with collective groups of beneficiaries and it is common for ethical trading CS to target groups or firms rather than individual producers. Therefore, the review considered as participants

agricultural producers and wage workers, whether individually certified or organised in POs. Households are also considered insofar as their members are part of a certified organisation or individually certified, or employed by a certified producers, with implications for the overall welfare of their households. Export firms or trading companies were not part of the participants. There are studies that have looked at effects of CS from the point of view of participating export firms but these were outside the scope of the review as they contained no evidence on primary agricultural producers or wage workers employed in agricultural production (e.g. Schuster and Maertens, 2013).

Studies also report on community-level impacts, especially in the use of the premium for social investments, as in the case of Fairtrade. Rijsbergen et al (2016) for instance, consider impact at a variety of 'levels', including the wider local community. Aggregate communities are not considered 'participants' strictly speaking, but the benefits of community interventions are considered in relation to individuals and households who are part of those communities. It is in this regard that issues of distributional dynamics matter most because communities are not simply homogenous groups of producers, workers or households, but are characterised by intrinsic inequalities that make aggregate community-level impacts hard to interpret.

The approach was to consider different potential participants' characteristics which could be used as potential moderators, such as age, gender, socio-economic status, location, type of production (smallholder vs plantation), type of product, types of certification scheme, and length of participation in the supply chain of the relevant agricultural certification schemes. As will be show in Section 4 reporting on some of these characteristics (e.g. age) is unusual besides broad descriptive statistics for samples of certified vs non-certified groups, and moderator analysis is constrained by data availability.

Despite the importance of CS for consumers and the fact that many standards and interventions are designed to directly and indirectly target and benefit consumers, this review only focused on the producer side. The ToC for consumers is too different and would warrant a separate review.

3.1.3 Types of interventions

This review includes studies on the effects of farm-level interventions in the production of agricultural commodities under certification schemes that have clearly defined socio-economic goals and third party auditing, even if socio-economic improvements are not the primary aim of the certification scheme. The certification schemes, such as interventions that follow the fair trade principles, as defined by the World Fair trade Organisation (WFTO), as well as other for examples under the social sustainability umbrella, must aim directly and explicitly to improve the wellbeing of beneficiaries. Farm-level interventions typically include the provision of training and expertise for improved agricultural practices, product marketing and/or self-organisation. CS may have additional community-level interventions, such as the provision of a price premium that is used by collective groups (such as producer organisations and cooperatives) to fund projects designed to benefit the group or the wider community in which they are located. The act of certification itself often entails

requirements that lead producers to invest and change practices. It is, therefore, *strictu senso*, a farm intervention which comes with auditing, checks and often training.

Interventions which simply aim at advocating the objectives and activities of, for example, fair trade or other forms of ethical trade were excluded, as they are designed to raise awareness among consumers without directly affecting the welfare of agricultural producers and workers.

Interventions and certification for the use of environmentally friendly production processes or environmental sustainability were also excluded, unless socio-economic outcomes are reported and/or the certification includes ethical trade standards in addition to environmental standards. There is a growing appreciation of the intertwined nature of social and environmental change processes, and the examples mentioned above attest to this reality. However, it is also true that certification schemes may aim to achieve environmental outcomes in their own right and with no necessary link with socio-economic outcomes. Previous (non-systematic) literature reviews (Chan & Pound, 2009; Nelson & Pound, 2009) have also noted the difficulties inherent in comparing and aggregating impact findings from studies focused on ethical trading and those dealing with environmentally-driven standards.

There are certification schemes, such as Rainforest Alliance, that have environmental sustainability as a primary outcome, but also have explicit objectives in relation to improvements in labour standards. Therefore, studies that include evidence of the impacts of Rainforest Alliance, or similar schemes, on their intended labour standards are screened for inclusion. Organic certification is a special case that had to be reconsidered in the course of the screening process. Generally organic standards focus on environmental sustainability and organic production practices, but there is substantial diversity especially if ‘organic by-default’ is included in the group, and some organic certifications also incorporate social sustainability (or ethical trade) standards that are directly relevant to socio-economic outcomes (Bennett and Franzel, 2013). Organic certification interventions are also substantially different as they produce a materially different product (one which is free from chemical residuals), while CS do not necessarily do so, as it is only the process of production which is supposed to be different (in terms of social and environmental sustainability). Moreover, the impact of organic certification systems on producer welfare deserves its own focused review and the general inclusion of the large literature around such certification would have expanded the boundaries of this review beyond the manageable, while unnecessarily diluting the conceptual boundaries of the research question pursued. Organic certification is however frequently used in conjunction with other certification system, and we have included studies reporting on such double, or even triple, certification.

Finally, to ensure a level of comparability between the CS examined, and as argued in Section 1, CS that are not third-party certifications, such as certifications internal to particular corporations (for example, Nestle’s AAA standard), were excluded. There may be a case for a review of evidence of labels and standards adopted as part of

firms' corporate social responsibility strategies, but they are intrinsically different in terms of origins, logic and governance to the CS reviewed here.

Section 1.3 discussed the range of interventions and the causal pathways in some detail. The interventions were classified in four blocks:

1. capacity building actions;
2. price and market interventions;
3. premium-funded social, community and business investments
4. labour standards monitoring

Within these blocks several interventions are considered. Below is a non-exhaustive list of interventions and components of certification schemes which were considered for the purposes of this review:

- Price floor (guaranteed minimum price);
- Price premium for individual producers;
- Credit and pre-payment;
- Longer-term and more stable contracts (market/trade relations);
- Development of standards requirements and auditing processes for compliance;
- Market access interventions through labelling and traceability;
- Provision of technical assistance and various forms of training for better farming practices, designed to increase the quality and productivity of farms, to meet more demanding market standards;
- Provision of technical or organisational assistance for POs or workers' organisations. Such interventions may include capacity building for farmers or workers, or thye may focus on improvements in quality, marketing and storage improvements, the development of record-keeping and management plans, as well as support for more effective and democratic governance of POs and plantations. The latter are meant to inform community-led decision making on investments and PO management, and prevent discrimination against vulnerable social groups, as determined by local context;
- Social or economic premium interventions that pay a premium for social or economic development projects which can be invested to improve production, marketing and/or community services and infrastructure under the assumption of widely shared benefits at community level.
- Labour standards interventions, which set standards for (minimum/living) wages and improved working conditions. Such interventions include the monitoring of workers' rights and labour standards violations, and educational activities on workers' rights and labour standards.

Table 1 below lists all certification schemes covered in this review. This list is not the result of a pre-selection of schemes beyond the key parameter set out above and in chapter one. Schemes are included in the review because we found includable studies that examine these reviews. The final list of included reviews is therefore driven by the available evidence, and no schemes that met the broad conceptual criteria already mentioned were excluded.

Table 1: List of certification schemes in included studies[41]

Review question	Included individual certification schemes
RQ1	4C, Fairtrade, GlobalGAP, Rainforest Allicance, Utz Of these, all apart from 4C appear in the meta-analysis.
RQ2	Better Cotton Initiative, Bird friendly, CAFÉ Practices, ETI, Fairtrade, GlobalGAP, Rainforest Alliance, Roundtable for Sustainable Palm Oil, Shop for Change, Utz

3.1.4 Types of outcome measures

The review includes studies that contain data on outcomes related to relevant theories of change. Outcomes may be intermediate or endpoint, intended or unintended. As discussed in Section 1, studies on CS frequently report on multiple socio-economic outcomes. Some are inter-related and can be grouped together and some are distinct.

The focus of the review is on the endpoint outcomes for wellbeing and empowerment of beneficiaries and the conditions of their activities. In other words, our *primary outcomes* are endpoint income and other social outcomes, including empowerment, (even if hard to measure), as well as key intermediate outcomes that are very important insofar as their links with interventions are expected to be more direct. For example wages and (net) returns to certified production are intermediate outcomes but strongly linked to endpoint primary outcomes such as household income. To be included for RQ1 studies had to report on at least one primary outcome.

The list of primary, endpoint and intermediate, and secondary outcomes is the following. This is an exhaustive list of all key relevant outcomes. However, due to data constraints, the fact that most studies only reported on some of these outcomes and that the indicators used were sometimes not sufficiently reliable, the meta-analysis was limited to some of the indicators listed below. The protocol also did not include behavioural outcomes (e.g. attitudes to investment patterns, farming practices), as we expected very limited evidence on these, and the review was focused on outcomes closer to impact along the causal chain. The indicators we could find data for are discussed in the results section.

Primary outcomes, divided into *endpoint* and *intermediate* outcomes, include:

1. Household income or consumption or other measure of socio-economic status (monetary measures of total household income or consumption, asset or wealth index, as used in Demographic Health Surveys for instance) (endpoint outcome).

[41] As discussed in Section 1, own-company standards such as Café Practices are outside the scope of the review. Café Practices is included in this listing as one included study (Heller, 2010) deals with this standard, alongside other certifications. Heller (2010) was included not for Café Practices but due to its findings on Fairtrade, organic and Utz certifications.

2. Health and education of adults and children (years of schooling, literacy, current enrolment status, work days lost due to illness, infant mortality rate) (endpoint outcome).
3. Gender equity in the outcomes above (*endpoint outcome*).
4. Producers' and workers' empowerment (*endpoint outcome*). In general it is hard to find studies that produce consistent measures of 'empowerment' and some of them overlap with outcomes mentioned above. There is a rich literature on 'women's empowerment' that helped operationalise this set of outcomes. Kabeer (2001, p. 81) broadly defines it as the "expansion in the range of potential choices available to women". However, there is a potentially a wide range of measures that attempt to capture effects of an intervention on empowerment. Some measures or understandings of 'empowerment' may be in the form of concrete outcomes such as the co-ownership of processing/trading businesses as in the case of Kuapa Kokoo in Ghana and Divine Chocolate for instance (Doherty & Tranchell, 2005), while others may be reported as subjective assessments (perceptions) of greater capacity to control and/or influence, change or participate in a value chain (for producers), or perceptions of greater capacity to engage in collective action for better working conditions in the case of wage workers. No studies that met the inclusion criteria for RQ1 reported directly on empowerment using reliable quantitative measures. Studies in RQ2 include some evidence on empowerment, but this could not be used for the purposes of statistical meta-analysis.
5. Gross or net returns to certified production (*intermediate outcome* as all other outcomes below), measured by different studies as gross/net farm profits or as farm revenue associated with a target crop.
6. Productivity of commodities (yield, that is, output per land unit, or output per worker or labour unit, if reported).
7. Price levels (for certified commodity and as farm-gate prices, that is, those effectively received by certified producers).
8. Wages (nominal and/or real, daily equivalent or other time unit). This outcome is of course part of household income, or contributes to it as an intermediate outcome, but may be reported separately as labour standards are a core component of many CS in social sustainability standards, so it should be assessed separately.
9. Non-wage labour conditions (health and safety: number of work-related injuries, access to health care, type of heath care available; benefits and entitlements: sick pay, paid holidays, maternity and paternity leave, free or subsidized food, clothing or shelter, freedom of association, and others).
10. Organisational empowerment of producers' and workers' organisations (that is, empowerment as a collective group and not just at individual level), which requires a consideration of the challenges in measuring empowerment as noted above (in order to operationalise, studies report various measures of enhanced capacity to benefit from the value chain or engage in collective action; this can take the form of direct participation in market institution

decision-making bodies or on concrete facts about successful collective bargaining).

11. Investments in services and infrastructure, funded by price premiums, as advances or direct transfers from certifying organisations.

Secondary outcomes include *both* endpoint outcomes (that are related to empowerment or equity) and intermediate outcomes, as follows:

12. Unintended outcomes (may be positive or negative/adverse); Unintended effects of certification, which can affect the above endpoint outcomes, such as effects on production costs (certification costs), debt, and workload, and local market conditions (that is, local prices, access to local markets) were also considered and analysed as part of the qualitative synthesis, in order to explore barriers and facilitators and how context creates conditions for some unintended effects.

To be included in the meta-analysis studies had to feature a controlled comparison between at least one group of certified farmers or workers and at least one comparable control group. The control group should be uncertified at the time of the study, though it is acceptable for the control group to be actively seeking certification itself. Moreover, we included studies that compare certified and cooperatively organised farmers with ones who are organised in a cooperative, but have no certification, and with farmers who are not members of any cooperative. In all cases treatment and comparison groups should be as comparable as possible in all aspects apart from certification, and studies should make efforts to match groups or control for group differences, or both.

3.1.5 Types of settings

The review considered evidence for LMICs. The focus is therefore on developing countries and on certain population groups within these settings.

Rural areas in LMICs where certification schemes are present are heterogeneous across contexts but there are some broad features in terms of agricultural production systems. These are summarized below:

- A prevalence of small-scale farms with production systems that combine food crop production for own consumption, with production of export or market-oriented crops.
- The presence of large-scale (in terms of capital invested) agribusiness in labour-intensive crops, as in the case of horticultural products, which has been expanding in the past thirty years, and has a substantial presence in Latin America and parts of Africa.
- In many of these settings production is organic by default, especially in Africa, because of constrained access to chemical inputs. However, there is great variation in the use of chemical inputs across countries and crops.
- There is a variety of business models for the organization of small-scale producers, but POs such as cooperatives have expanded across countries in response to the gaps left by market liberalization, and contract farming has also become more common in many LMICs since the 1990s (Oya 2012).
- These are also settings in which rural poverty is widespread and where small

producers and agricultural wage workers are over-represented among the poor. However, there is substantial differentiation among smallholder producers in terms of poverty levels.

3.2 Search methods for identification of studies

3.2.1 Electronic searches

We searched for studies that report on the effects of certification schemes (CS) for agricultural commodities and their associated interventions, as well as for studies that examine the circumstances under – and the reason(s) for – which such interventions have intended or unintended effects (i.e. barriers and facilitators to CS effectiveness). The nature of the literature on this topic and its heterogeneity required a multi-pronged strategy to find relevant evidence, whether from independent academic studies published in peer-reviewed journals, or from impact evaluations and studies commissioned by certifying bodies or funders of CS. In addition to searching academic databases, we therefore conducted extensive targeted searches for reports and papers not indexed in databases. In accordance with guidelines by Hammerstrøm et al (2010) for Campbell Systematic Reviews, we worked closely with information retrieval specialists to help devise and quality-assure our general search strategy, and to ensure that it is as exhaustive as possible. Consequently, two information retrieval specialists helped develop search strings and provided feedback on troubleshooting certain problematic databases (in particular AGRIS, Agricola, IBSS, Africa Wide, JOLIS, Labordoc, AgEcon, IDEAS, International Institute for Environment and Development) as well as on the best ways of conducting targeted searches.

In order to produce a comprehensive list of keywords related to the review's inclusion criteria (PICOs), we combined brainstorming and pearl-harvesting (collecting keywords from studies that meet the inclusion criteria) as suggested by Sandieson (2006). A single general search strategy was used to find both quantitative and qualitative studies for RQ1 and RQ2, however, this was adapted to each database and website searched. For each database, we studied its thesaurus and customised our general strategy accordingly, including the appropriate vocabulary. For websites, which tend to have much more limited searching facilities, multiple targeted searches were conducted using selected key words from the general search strategy. Our basic search strategy was adapted to each database, combining text terms with indexing terms using Boolean (AND /OR) and proximity (NEAR/WITHIN/ADJ) operators: '[population terms] AND [intervention terms] '.

All customised search strategies were piloted in order to assess their relevance and precision, and to identify the most optimal set of search terms. We prioritised high sensitivity of the search terms over precision in order to avoid omitting relevant studies which do not report sufficient information in their title or abstract. Reviewers were over-inclusive during the first stages of screening titles and abstracts.

We used the EPPI-Reviewer 4 bibliographic software (Thomas et al, 2010) to manage retrieved references. Where possible all references were downloaded along with the necessary fields (i.e. abstract, article identifier, index terms/thesaurus) and

imported into EPPI-Reviewer 4. Where the reference source did not support export facilities, relevant references were imported manually. As explained below, duplicates were removed automatically with EPPI-Reviewer 4. In addition, a large number of duplicates were removed manually during the screening process. This was because the reporting of titles and even authors or years was not always consistent, thus often there were no 'perfect' duplicates, which delayed the process of duplicate removal. All the searches, including hand searches, were documented and a detailed record of the type of search (i.e. electronic, hand-searching, etc), specific search strategy, number of references retrieved, date of search and search source is available upon request. For the main database searches, the examples of the search strings and dates are given in Annex A, as is an example of our hand-searching strings and search methods.

To ensure we conduct the most comprehensive search possible, we searched multiple databases, as suggested by Hammerstrøm et al (2010), including general social science-related bibliographic databases, subject-specific databases covering agriculture and international trade/economics, systematic review databases, and national and regional databases. The main database searches were conducted between May and October 2015. The last main database was searched on 10 October 2015. For all searches we show the date of the last search and the main website in brackets. We covered the following databases:

- AgEcon, 17/07/2015 (http://ageconsearch.umn.edu/)
- Africa Wide, 07/05/2015 (https://www.ebscohost.com/academic/africa-wide-information)
- CAB Abstracts, 05/07/2015 (http://www.cabi.org)
- International Bibliography of the Social Sciences (IBSS), 07/05/2015 (http://www.proquest.com/libraries/academic/databases/ibss-set-c.html)
- Social Sciences Citation Index (SSCI) / Web of Science, 02/05/2015 (http://isiknowledge.com)
- Econlit, 06/05/2015 (https://www.ebscohost.com/academic/econlit)
- US National Agricultural Library, 12/08/2015 / AGRICOLA, 01/10/2015 (http://naldc.nal.usda.gov/naldc/home.xhtml / http://agricola.nal.usda.gov/)
- JOLIS (WB/IMF), 16/09/2015 (http://external.worldbankimflib.org/external.htm)
- British Library for Development Studies (BLDS), 21/07/2015 (http://blds.ids.ac.uk/)
- IDEAS repec, 10/10/2015 (https://ideas.repec.org/)
- International Institute for Environment and Development, 12/09/2015 (http://www.iied.org/)
- 3ie systematic reviews and impact evaluations database, 03/08/2015 (http://www.3ieimpact.org/en/)
- ILO Labordoc, 11/09/2015 (http://labordoc.ilo.org/?ln=en)
- The Campbell Library, 18/08/2015 (http://www.campbellcollaboration.org/lib/)

We also searched grey literature databases, as well as websites of research institutions, organisations related to CSs for agricultural commodities, funders and

donors. Much of published material on certification schemes, especially those related to ethical trading, is produced by certification bodies, as well as by activists and NGOs seeking to further ethical trade. We therefore invested considerable time and effort in moving beyond the standard social science databases and including less commonly searched specialist databases, as well as potential sources of grey literature, such as the website of certification bodies, NGOs and research institutes. Several organisations also maintain large, but not public, archives of impact studies, but we were only successful in accessing these in a small number of cases.

In almost all cases the databases and websites searched during hand-searches do not support complex search strings or allow for the direct export of material, and searches are therefore much more 'manual' and time-consuming. Websites and specialist databases were searched in a flexible manner that combined the search process with screening for inclusion. The search process had to be flexible due to the wide variation in functionality across websites. Some have reasonably well-developed search functions, while others have to be browsed by keywords, or even browsed in their entirety. For each website, the search function was tested first using a list of relevant keywords. If the search function proved satisfactory in terms of returning potentially relevant material, then this function was used. If not, as was commonly the case, the website was browsed by keywords or content categories. If these were not available, the website's publications (or similar) sections were browsed. Potentially relevant studies were then screened via title and abstract and a partial review of the text. To allow searches to be replicated, all website navigation and all search terms were recorded for each website, along with the number of returned results and the outcomes of the screening process. Lists of all included studies from each website or database were maintained as part of the search documentation. Please see Annex A for some example search protocols. All hand searches took place between July and December 2015, and the final search was on 26 December 2015. We covered the following specialist databases and websites:

- Agra.org, 18/10/2015 (http://www.agra.org/)
- AGRIS / KOHA (FAO), 15/11/2015 (FOA/Koha), 28/12/2015 (main FAO site) (http://agris.fao.org/agris-search/index.do / http://unfao.koha-ptfs.eu/cgi-bin/koha/opac-search.pl / http://www.fao.org/home/en/)
- Canaan, 24/08/2015 (https://www.canaanusa.com)
- Catholic Relief Services Fair Trade, 24/08/2015 (http://www.crsfairtrade.org/)
- Centre for Fair and Alternative Trade, 05/09/2015 (http://cfat.colostate.edu)
- CGIAR, 24/12/2015 (http://www.cgiar.org/)
- COSA, 24/08/2015 (http://thecosa.org)
- ELDIS/Institute of Development Studies (IDS), 22/12/2015 (http://www.eldis.org/)
- ESRC (Economic and Social Research Council), 30/11/2015 (http://www.esrc.ac.uk/)
- European Fair Trade Association, 04/10/2015 (http://www.european-fair-trade-association.org/efta/index.php)
- Fair Trade Institute, 23/07/2015 (http://www.fairtrade-institute.org/)

- Fair Trade USA, 27/08/2015 (http://fairtradeusa.org/)
- Fairtrade Foundation, 20/08/2015 (http://www.fairtrade.org.uk/)
- Fairtrade International, 04/10/2015 (http://www.fairtrade.net)
- GlobalGAP, 27/08/2015 (http://www.globalgap.org/)
- IFPRI, 28/08/2015 (http://www.ifpri.org/)
- ISEAL Alliance, 20/10/2015 (ISEAL sent to call for papers to all members on our behalf)
- MPS (Fair flowers fair plants), 25/08/2015 (http://www.fairflowersfairplants.com/home-en.aspx / http://www.ecas.nl/en/
- Networked Digital Library of Theses and Dissertations, 26/12/2015 (http://www.ndltd.org/)
- Oxfam, 25/08/2015 (https://www.oxfam.org/)
- ProQuest dissertation database, 12/12/2015 (http://search.proquest.com/pqdtglobal)
- R4D, DFID, 02/10/2015 (http://r4d.dfid.gov.uk/)
- Rainforest Alliance, 26/08/2015 (http://www.rainforest-alliance.org/)
- Soil Association certification (ethical trading), 26/08/2015 (http://www.soilassociation.org)
- Traidcraft, 25/08/2015 (http://www.traidcraft.co.uk/)
- TWIN, 24/12/2015 (www.twin.org.uk)
- USAID, 19/09/2015 (https://dec.usaid.gov/dec/home/Default.aspx)
- Utz Certified, 18/11/2015 (Utz provided material to us directly)
- Wageningen University and UR, 23/12/2015 (http://www.wageningenur.nl/en/)
- World Fair Trade Organisation, 09/09/2015 (http://wfto.com/)

Finally, we covered relevant websites and databases of studies German, Portuguese and in Spanish. This required the targeted selection of relevant databases, especially for Latin America (such as Scielo) and websites associated with evaluations of interventions in agricultural production, including standards, in non-English language.[42] We did not search specific French-language databases but did consider French-language material discovered during our searches for inclusion. We searched the following non-English language databases:

- SCIELO [Portuguese & Spanish], 03/12/2015 (http://www.scielo.org/php/index.php)
- GESIS SOWPORT [German], 19/11/2015 (http://sowiport.gesis.org)

With Germany being a leading market for ethically traded products a particular emphasis was placed on finding German-language material.[43] The following German-language websites were searched:

- Brot für die Welt, 22/11/2015 (http://www.brot-fuer-die-welt.de)
- Bundesministerium für wirtschaftliche Zusammenarbeit (BMZ), 18/11/2015 (http://www.bmz.de)

[42] For example a centre for the study of coffee in Colombia where a number of evaluations were conducted, including on voluntary standards http://www.crece.org.co/crece/
[43] These searches however identified little or no useable material.

- DEval – German Institute for Development Evaluation, 17/11/2015 (http://www.deval.org)
- Fairtrade Deutschland, 18/11/2015 (https://www.fairtrade-deutschland.de)
- Forum Fairer Handel, 20/11/2015 (http://www.forum-fairer-handel.de)
- GEPA, 19/11/2015 (http://www.gepa.de/home.html)
- Misereor Deutschland, 18/11/2015 (https://www.misereor.de)
- Südwind Institut, 20/11/2015 (http://www.suedwind-institut.de/)

The following databases, which we had committed to searching in the protocol were found to be irrelevant or inaccessible during the search process and hence were dropped:

- Best Evidence Encyclopaedia (BEE) (irrelevant material)
- Fair Trade Resource Network (the site closed, material was migrated to the Fair Trade Institute, which was searched)
- Dritte Welt Laden (no substantial material)
- World Bank (as the JOLIS database contains all WB material)

3.2.2 Searching other resources

We used snowballing on a continuous basis while "the study unfolds", as recommended by Greenhalgh and Peacock (2005, p. 1064). In order to locate eligible studies for RQ1 that could have escaped the electronic search we screened the references of included studies and conducted a systematic reference checking and citation tracking of studies included for RQ1 in February 2016. Given the large amount of studies eligible for RQ2 and the limited resources of this review the same was not possible for studies included for RQ2.

We also used our existing knowledge of the literature and our networks, including the advisory board, and organisations working in the field of CS in order to identify additional eligible studies, including unpublished papers or on-going research. We asked leading certification bodies and research institutions to provide lists of impact studies they were aware of and several responded. Finally, we shared a preliminary list of included studies with ISEAL, the umbrella organisation for many certification bodies for them to circulate amongst their members and suggest studies we might have overlooked. This allowed for additions of very relevant studies published after our searching process was completed (i.e. Rijn et al, 2016; Waarts et al, 2016; Aidenvironment, 2016). Last but not least, we were "alert to serendipitous discoveries", i.e. finding a relevant study when looking for something else (Greenhalgh and Peacock, 2005, p. 1065).

3.3 Data collection and analysis

3.3.1 Selection of studies

Owing to the range of interventions assessed and the nature of the literature – which as discussed spans both academic and non-academic publications – the study selection process had to be able to accommodate and evaluate a substantial variety of study types and designs, as well as project reports and advocacy publications. As discussed in the search section above, studies were retrieved from a wide array of

databases and websites. The sheer number of retrieved results necessitated a multi-stage selection process. At each stage, studies were assessed against clear sets of inclusion criteria and studies that did not meet the criteria were excluded. Studies were managed and coded for inclusion and exclusion using EPPI-Reviewer 4. All coding decisions were retained in the software. The coding tools for each stage were piloted until a sufficient degree of inter-coder agreement was achievement. All coders were over-inclusive, so as not to lose potentially relevant material. Where coding was undertaken by research assistants, they were given extensive training, were closely supervised and were given detailed written coding instructions. In total, we undertook two stages of single screening titles and abstract, one stage of single screening using a partial text review, and then double screened on full texts. Even then the number of studies for RQ2 was excessive given our resources, so we undertook another stage of single full-text screening for RQ2 only using a more restrictive set of inclusion criteria, conducted by the core review team. Studies at full text stage were double coded independently by two coders, with disagreements between coders resolved by the PI. The coding at earlier stages was reviewed by DS, CM, FS and the PI to ensure inclusion and exclusion decisions complied with the relevant criteria. The numbers of studies identified, included and excluded at each stage are discussed in Section 4 below. Duplicates not identified at initial stages were removed during the initial screening stage. A second round of duplicate removal, this time without using EPPI-Reviewer's automated identification system, was undertaken prior to the partial text review screening stage.

The first stage of screening on titles and abstracts was conducted by three research assistants overseen by DS and CO. Studies were screened for inclusion against basic relevance criteria developed from our PICOs. Studies imported form databases were screened in EPPI-Reviewer 4. As exporting search results was not possible for the majority of databases and websites covered during hand-searching, these results were screened directly during the search process. All of these latter searches and coding decisions were recorded in spreadsheets. Studies coded for inclusion during the hand-searching phase were then manually imported into EPPI-Reviewer 4.

Due to the often limited amount of information conveyed by titles and abstracts, large numbers of irrelevant and ineligible studies were still included after the first round of screening. A second round of screening on titles and abstracts was therefore undertaken to identify and remove material that did not contain primary evidence. In particular, this round targeted literature reviews and articles not based on empirical research. Coding was undertaken by four research assistant overseen by DS and FS.

Despite this second round of screening on titles and abstracts, the number of items included at this stage was still too large to be screened on full texts, given the resources available to us. We therefore decided to engage in a partial text review aimed at excluding studies not based on primary data or new analyses of existing data. Such studies can rarely be identified from title and abstract alone, but can be found quickly by a targeted review of the main text body.

Left with only material based on empirical studies we then engaged in full text screening. Separate coding tools were developed for RQ1 and RQ2 to reflect the different criteria of inclusion. Both tools were extensively piloted to ensure agreement between coders. Studies were then double-coded independently by two coders. Disagreements in coding were traced in EPPI-Reviewer 4 and were mediated by the PI. There were more disagreements for RQ2, as the criteria for inclusion in that review question necessarily had to rely less on well-defined aspects of study design and analytical method. Non-English studies had to be single coded due to resource limitations. Studies in French and Portuguese were coded CO, studies in Spanish and French by DS and studies in German by FS.

After full text screening the number of studies included for RQ2 was still too large to permit a meaningful synthesis. We therefore decided to engage in another round of full text screening for RQ2, using a tighter set of inclusion criteria focused on relevance to three thematic areas namely implementation dynamics, distributional dynamics and causal mechanisms to impact. As studies were single coded at this stage the coding tool was piloted extensively and the coding was reviewed and arbitrated by the PI. The coding was carried out by DS, CM and the PI.

3.3.2 Data extraction and management

Once the studies to be included for RQ1 and RQ2 had been decided, data extraction proceeded separately for studies eligible for RQ1 and RQ2. Studies eligible for inclusion for both RQ1 and RQ2 underwent both processes of data extraction. For both review questions data were extracted using detailed coding sheets in MS Excel; for RQ2 additional data extraction was untaken in NVivo. For RQ1 the relevant data were analysed using Stata 14.2, while for RQ2 data analysis was undertaken using both MS Excel and NVivo. Coding sheets for both questions were developed by CO, DS, LL, CM and FS. Both data extraction sheets were carefully trialled prior to use to ensure both coder agreement and efficiency.

For RQ1, data were extracted by FS and LL and the results were audited by the lead reviewer, CO. Effect size calculations were undertaken by LL, with additional support provided by staff at 3ie, and were reviewed by CO. LL also lead on the risk of bias assessments for RQ1, with additional assessments made by FS and the results reviewed by CO. Data extraction for RQ2 was conducted by DS and CM, with oversight from CO. Reviewers searched for relevant data in the entire text of included studies and not only in specific study sections. The study quality assessment for studies included in RQ2 was conducted by DS, CM and CO.

Coding sheets for both review questions extracted basic bibliographic and contextual information from the included studies. Overviews of coding sheets for both review questions can be found in Annex D. In addition, the RQ1 coding sheet focused on capturing the data needed to calculate effect sizes, as well as possible moderators, and the information necessary to conduct sensitivity assessments, including risk of bias. By contrast, the RQ2 coding sheet collected information related to the implementation of schemes, distributional dynamics among scheme participants, and contextual barriers and facilitators, especially as these relate to achieving stated goals, as well as the information on methodological aspects of the studies, necessary

to assess study quality. We remind the reader that RQ2 was not concerned with the impact of CS (i.e. whether CS work or not), as in the case of RQ1, but with the barriers and facilitators mediating the effects of CS (i.e. how, why, when, for whom CS may or may not work). Therefore, the coding sheet developed for qualitative data extraction sought to extract data that could help illustrate these aspects of CS. As a result, descriptive data from qualitative studies that only reported on CS outputs, without providing any insights on implementation or distributional dynamics, or on how the context can shape these outputs, were excluded. For instance, simple descriptions of the use of social premium, which were commonly found, were excluded from data extraction, unless they were accompanied with some insights on how related decisions on the premium investments were reached, who was able to benefit from these investments (or who was not, and why), how the context played a role in the success or failure of such investments, etc.

Furthermore, during the data extraction stage for RQ2 we sought and coded for text that presented substantive and relevant evidence linked to primary data, i.e. descriptions of how certification programmes were implemented, how certified cooperative or plantations operated, etc. Authors' statements or opinions that were not supported by presented findings or context descriptions were not coded for synthesis. Similarly, qualitatively-researched perceptions of CS effects (i.e. farmers' perceptions on the benefits resulting from participation in CS) were not included unless accompanied with factual descriptions or explanations of how or why these perceptions were formed.

To further illustrate this distinction, we provide an example of excluded and included extracts from the same study in Annex E.

3.3.3 Assessment of risk of bias in included studies

The risk of bias assessment and critical appraisal differed for studies included for RQ1 and RQ2, as many of the criteria used to assess study quality in quantitative impact studies are not meaningful in assessing qualitative work. To assess the impact evaluation studies included for RQ1 we built on the bias assessment tool developed by Waddington et al (2014), which categorizes studies as either low, medium or high risk of bias. As laid out in our study protocol (Oya et al, 2015), we assessed the studies for RQ1 across the following seven dimensions:

- Selection and confounding: Was the allocation or identification mechanism able to control for selection bias and confounding?
- Group equivalence: Was the method of analysis executed adequately to ensure comparability of treatment and control groups throughout the study and prevent confounding?
- Motivation bias: Was the study adequately protected against the process of being observed causing motivation bias? (These possible biases are known Hawthorne and John Henry effects.)
- Spill-over effects: Was the study adequately protected against performance bias, in that treatment and control groups were geographically and/or socially separated?
- Selective reporting of outcome: Is the study free from evidence of the

selective reporting of outcomes (as far as this can be assessed given the lack of published pre-analysis plans in the assessed literature).

- Selective analysis: Did the authors use 'common' methods of analysis and is the study free from the suggestion of biased exploratory research methods (again, there are limits to how far as can be assessed given the lack of published pre-analysis plans and replication)?
- Other bias: Was the study free from other form of bias, including, but not limited to, courtesy bias from self-reporting, the retrospective collection of 'baseline' data, the validity of instruments and other issues with measurement?

For each question coders we asked to answer either 'yes', 'unclear', or 'no' for each study. Where a study did not report the necessary information, it had to be coded as 'unclear'. Inadequate reporting practices therefore increased the risk of bias ratings for many studies. The overall risk of bias score for RQ1 studies was arrived at by aggregating the answers to each of the questions above. As selection bias is the most serious methodological issue affecting impact studies, and especially so in the field of certification schemes where self-selection into certification systems is the norm, we gave a greater weight to methodological weaknesses is this area.

Accordingly, studies were assigned a low risk of bias rating if all questions were answered with either 'unclear' or 'yes' and scored at least one 'yes' for selection or group equivalence. Studies were given a medium risk of bias rating if they scored 'unclear' for both selection and group equivalence and no worse than 'unclear' in all other categories. A high risk of bias was given to studies if they either scored 'no' for either selection or group equivalence and 'no' in one other category. Finally, studies were rated as critical is they scored 'no' for both selection and group equivalence and also displayed two other forms of bias. Studies rated as critical risk of bias were retained in the review but were excluded from the meta-analysis, and their exclusion is noted in the relevant sections below. The full risk of bias assessments for all included studies are available in Annex C.

Four of the studies study included in the review, all reported in Cramer et al (2014), were co-authored by the review's principal investigator. To prevent any conflict of interest, the risk of bias assessments for these studies were undertaken solely by, LL, who joined the team as a consultant and has not had any previous association with any of the authors of Cramer et al (2014).

As discussed in our protocol (Oya et al, 2015), unlike for quantitative impact assessments, there is no generally accepted standard for assessing the methodological rigour of qualitative studies (see also Vaessen et al, 2012). Moreover, the role of the method assessment for RQ2 is very different than for RQ1. Where the risk of bias assessment for RQ1 is an important aspect of conducting sensitivity analysis of the results of the impact synthesis, the quality assessment for RQ2 serves primarily to assure that only studies of a minimal acceptable quality are synthesised, as studies that lacked reporting of basic methodological aspects were excluded, as well as to provide information on the quality of evidence used in the

synthesis for RQ2. Part of the quality assessment for RQ2 studies was therefore already undertaken during the full text screening stages (see Section 3.1.1 on methodological inclusion criteria for RQ2). At this stage, the quality assessment was not used to further exclude studies, but to provide an overall picture of the quality of the studies included.

For the quality assessment for RQ2 we adapted the CASP (2006) tool and the approach used in Waddington et al (2014) to our needs. Our assessment combined both substantive and reporting-based criteria. Study quality was assessed as either high confidence or low confidence along the following dimensions:

- Clarity of research question
- Justification of research approach and its appropriateness to the research question
- Clear description of study context
- Clear description of the researcher's or researchers' role
- Where appropriate, clear description of the sampling methods used, and the suitability of the sampling strategy
- Justification of the selection of the research site or sites and an assessment of whether the choice is appropriate
- Clear description of the methods used to collect data and an assessment of whether the methods are appropriate to answering the research question
- Clear description of the method or methods used for data analysis and whether this is deemed appropriate
- Assessment of whether the data collected supports the claims made by the study
- Assessment of whether data are triangulated, that is, cross-verified across two or more sources

The assessment of study quality for RQ2 is complicated by the fact that qualitative research, by its very nature, covers a much wider array of methodological approaches that quantitative impact assessments do. In particular, we undertook special efforts to include high-quality ethnographic studies, which benefit from long-term fieldwork and are frequently published as monographs, meaning they can go into considerable detail in discussing how certification 'works' on the ground. Ethnographic studies, however, are based in a methodological approach that does not use concepts such sampling, making part of our quality criteria ill-suited to the assessment of ethnographic material. Unlike for RQ1, we therefore refrained from producing a synthesised quality assessment metric for RQ2 studies, as any such metric would have been largely driven by its method of construction and there would have conveyed little information to the reader. Instead we have published the full quality assessment for each included study for RQ2 in Annex C.

3.3.4 Measures of treatment effect

In all studies the treatment variables attempted to measure the impact of a certification scheme on an outcome – or set of outcomes – of interest. As laid out above, a certification scheme comprises a bundle of interventions; normally at least

the process of certification itself as well as some form of training or advice to help achieve the required standards. The bundles represented by the different CS are therefore quite different. To complicate matters, CS are frequently not the only development intervention in a given area. Some NGOs, for instance, provide training to producer organisations to help these groups achieve certified status, so as to support wider goals of community development (see, for example, Hoebink et al, 2014). In practical terms it is impossible to separate out fully the effects of certification from the effects of other development interventions, as the presence or absence of other interventions in or around treatment and/or control groups is not reported in many cases, and may even be unknown to the researchers themselves.

We calculate standardised mean differences (SMDs) for all studies for which we could extract the necessary data. The SMD, or Cohen's *d*, is defined as the unstandardized difference in means between the treatment and control groups divided by the pooled standard deviation of the outcome measure, i.e. the standard deviation across both treatment and control groups:

$$SMD = \frac{Y_t - Y_c}{S_p}$$

$$SMD = t \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Where Y_t is the mean of the treatment group and Y_c is the mean of the control group on the outcome measure of interest, and S_p is the pooled standard deviation. However, the pooled standard deviation of the outcome measure is not commonly reported and cannot always be calculated from reported data. In these cases it is possible to instead use the standard deviation of the outcome in the control group only, or to approximate the SMD using the formula below (Waddington, 2014):

Where *t* is the t-statistic of the between groups mean test or the t-statistic of the treatment regression and n_1 and n_2 are the sample sizes of the treatment and control group, respectively.

For small sample sizes Cohen's *d* is slightly positively biased and a small correction is necessary. We have adjusted our SMDs and their SEs using Hedges' method, thus converting Cohen's *d* into Hedges' *g*. The SMDs were converted using the formula given by Ellis (2010).

$$g \cong d\left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

SMDs and their standard errors were calculated using Wilson's web-based effect size calculator.[44] The calculator uses standard formulae provided in Lipsey & Wilson (2001).

[44] The calculator can be found at http://www.campbellcollaboration.org/this-is-a-web-based-effect-size-calculator/explore/this-is-a-web-based-effect-size-calculator.

To aid the interpretation of the results of our meta-analysis, we have translated the SMDs into 'standardised percentages' using binomial effect size displays (Randolph & Edmondson, 2006). The binomial effect size display is created by first translating SMDs into point-biserial correlation coefficients (*r*). Then success rates are calculated for treatment and control groups. The base success rate in both groups is arbitrarily set to 50% and deviations from this base rate are determined by the size of the correlation coefficient. In the executive summary we published the difference in percentage points between the calculated success rates for treatment and control groups. We also present upper and lower bounds for this difference, based on the inter-group differences at the upper and lower bounds of the correlation coefficient's confidence interval. Binomial effect size displays are however not equivalent to a translation of effect sizes into the units used by the underlying studies. They are not the same as a percentage change in the underlying raw data and should not be interpreted as such. Their sole purpose is to communicate effect sizes in a more intuitive format and for a clearer sense of orders of magnitude.

3.3.5 Unit of analysis issues

As a robustness check we provide our core results both with and without applying

$$SE_{Corrected} = SE_{Uncorrected} \times \sqrt{[1 + (m - 1) \times ICC]}$$

corrections for unit of analysis errors. Correcting for possible unit of analysis errors involves recalculating the standard errors around the effect size estimate. Standard errors, and therefore confidence intervals, are inflated to account for the fact that there might be correlation among the different clusters in a study. Based on Higgins and Green (2011) and Waddington et al (2014) we used the formula below to recalculate standard errors.

Where *m* is average cluster size (the number of respondents in each cluster, in our case) in the study and the *ICC* is the intra-cluster correlation coefficient, which we assume to be 0.05.

Not all studies provided the necessary information though. For 14 effect sizes across eight studies we had to impute information (Anteneh et al, 2014; Becchetti & Gianfreda, 2008; Becchetti et al, 2008; Bennett et al, 2012; Parvathi & Waibel, 2016; Riisgaard et al, 2009; Subervie & Vagneron, 2013; Weber, 2011). Given the uncertainty these assumptions adds to the calculations, we decided to present our results in the main body of the review without unit of analysis correction. Results to which the corrections have been applied are shown in Annex F. The differences between both sets of results are minor and are detailed in the annex.

3.3.6 Dealing with dependent effect sizes

Only effect sizes that are statistically independent are to be included in any one meta-analysis of a given outcome. Combining more than one effect size from a study in the meta-analysis of an outcome violates this assumption of independence. To identify and prevent the inclusion of dependent effect sizes we differentiate between a study, which we define as a unique dataset based on a particular sample, and a report, by which we mean a write-up of a study. So there may be multiple reports of a single study, either because of different publication formats (for example, a working

paper followed by a journal article), or because they report on different outcome measures or aspects of the study. A similar situation arises when multiple reports draw on the same dataset. Conversely, a report may contain data on outcome measures from several studies. For instance, Cramer et al (2014) contains wage data from four separate datasets, which are all considered separate studies. Dependent effect sizes occur when a single report includes the findings of more than one study, or – more commonly perhaps – when more than one report exists on the findings of a single study. They can also occur when studies include more than one intervention group compared to one control group, when studies report outcomes at more than one point in time, or where studies report multiple specifications, multiple outcome constructs (or groups thereof) or report results for different subgroups of participants.

Our coding scheme recorded information on the certification scheme, the intervention, the country and area, and the timeframe of the intervention. We therefore coded for key contextual information necessary to identify dependent effect sizes across and within studies, allowing us to combine reports of a single study as well as to separate out multiple studies contained in a single report prior to calculating effect sizes. Unique study- and report-level identifiers were used to track both studies and reports through the data extraction process. One case where studies overlapped in terms of their samples are van Rijsbergen et al (2016), Ruben et al (2014) and Kamau et al (2010), as these share control groups. In this case we elected to use the impact estimates presented in Rijsbergen et al (2016) which uses panel data and has the lowest risk of bias rating. Where studies reported multiple treatment arms, which can be the case then one treatment group is double certified for instance, we chose the treatment arm that was single certified.

We have therefore included only one effect size estimate per study in any given synthesis.[45] Where multiple outcome estimates were reported for the same outcome, commonly due to different model specifications in regression analysis or to the use of different matching algorithms, we selected the specification or method preferred by the study authors. If no preference was specified we chose the estimate likely to have the lowest risk of bias. We did not calculate any synthetic effect sizes.

3.3.7 Dealing with missing data

Due to time and resource constraints we were not able to contact authors of studies which did not provide sufficient information. For studies which met the inclusion criteria but did not report the information necessary for effect size calculations we have noted the direction and significance of their main findings in our results section. As these effect estimates have not been standardized we do not report the precise estimates, as these would be difficult to compare. The findings of these studies thus are reflected in our quantitative synthesis but did not contribute to our meta-analytic results.

[45] A synthesis refers to a meta-analysis of effect sizes on a given outcome. While only one effect size per study may enter into a given meta-analysis, studies may contribute effect sizes to more than one synthesis.

Wherever possible we tried to calculate effect sizes from the available information using approximations for the SMD, as provided by Waddington (2014), wherever necessary.

3.3.8 Assessment of heterogeneity

We provide forest plots and heterogeneity statistics for all outcomes for which we were able to synthesise data. In particular we report the heterogeneity Chi-squared and its associated p-value, as well as the T^2 and I^2 statistics. The I^2 statistic reflects the amount of variation in the finding that is due to real heterogeneity among studies, while the T^2 statistic is the estimate of the between studies variance component (Borenstein et al 2009).

At protocol stage, we had planned to conduct a substantial amount of moderator analysis (see Oya et al, 2015). This was however hampered by the limited amount of data we were finally able to extract. In practice the limited number of studies available for most outcomes meant that we only looked at the certification scheme type as a substantive moderator. While this moderator did therefore not arise ad hoc, the limitation of our moderator analysis was unforeseen. Data limitations also meant that we eschewed the use of meta-regression.

3.3.9 Assessment of reporting biases

We assess the potential of reporting and publication bias through sensitivity analysis of our meta-analytic findings and through the visual inspection of funnel plots (see Section 4.3). As the literature under review concerns a wide range of different interventions across a range of value chains, funnel plots are especially difficult to interpret in this instance. This problem is compounded by small number of available studies, which means we have limited the use of funnel plots to only those outcomes with the most available evidence.

3.3.10 Data synthesis

Review question (1)

Where the data allowed, we have synthesised the extracted effect sizes using inverse variance-weighted random effects meta-analysis (see also Lipsey and Wilson 2001). The meta-analysis was undertaken using the *metan* command in Stata 14.2. Random effects meta-analysis does not assume that there is a single 'true' effect size to be estimated across studies, but rather allows for a variety of underlying effect sizes. In the random effects model the total observed variance can be broken down into within-study variance and (estimates of) between-study variance. Given the real heterogeneity of study contexts and interventions in the literature under review, the random effects model appears appropriate. The inverse variance-weighted random effects model weights each study by the inverse of its variance (which captures sample size) and by an estimate of the between-studies variance component (T^2), thereby both giving greater weight to more precisely measured effect sizes and producing wider confidence intervals[46]. A fixed effect synthesis, by comparison,

[46] Using notation from Borenstein et al (2009) T^2 is the estimated value of the real between-study variance τ^2.

would produce artificially narrow confidence intervals, thus overstating the significance of effects (Borenstein et al 2009). As well as reporting the relevant synthesised effect sizes and diagnostic statistics, we display the results of our meta-analysis in detailed forest plots, complete with appropriate confidence intervals.

Review question (2)
Following Thomas and Harden (2008), we used a thematic synthesis approach in three stages to address the review question 2. As described above, we first developed a coding tool using the hypothesised programme theory focused on the key themes of interest for review question 2, i.e. implementation dynamics; distributional dynamics; contextual factors. Under each of the above themes a number of sub-themes were initially considered (for example, costs of certification, gender dynamics in distribution), but grounding further sub-themes in the process of coding data was also expected (see below). Every piece of relevant data could be coded under one or multiple codes, a fact which allowed us to observe how the different themes interrelate, which informed the structure of the final synthesis. Included studies were coded using both MS Excel and Nvivo.

Coding the studies for RQ2 was an iterative process. During the detailed coding, emerging themes not previously considered were identified and incorporated into the coding tool. Some were incorporated as new codes, and others were merged with already existing ones, creating new and more refined codes. The flexibility of Nvivo allowed us to develop a much more detailed, multi-layered hierarchical tree structure, while our spreadsheet coding tool remained simpler. Both tools, however, maintained the same basic structure. The respective final coding structures are provided in Annex D. In short, our themes were structured around our key predefined interests, and enriched and expanded during the coding process by new emerging themes grounded in the data in each study. No particular themes were excluded, unless they were deemed irrelevant to our three key themes.

After coding all the primary studies, we proceeded to the second stage, the generation of detailed descriptive themes (Thomas and Harden, 2008). First, we reviewed our coding structure for similarities and differences between the codes, and adjusted the structure by relocating and merging codes accordingly. Then, following Waddington et al (2014), we summarised the extracted data across studies under each code to produce a descriptive synthesis, which was still close to the extracted original text.

In the third stage, we built on the descriptive synthesis to generate "analytical themes". This was done by interpreting the descriptive summaries to provide "new interpretive constructs, explanations or hypotheses" (Thomas and Harden, 2008:n.a) that could inform our second review question. "Going beyond" the content of the primary studies can be the most controversial part of a qualitative synthesis, as Thomas and Harden underline, since it depends on "the judgement and insights of the reviewers". In order to reduce the potential influence of the researcher on the transition from descriptive to analytical themes, each analytical theme was discussed by the three reviewers (DS, CM and CO), who come from different disciplinary backgrounds but have extensive experience in qualitative data analysis. The

narrative synthesis therefore does not derive from the interpretation of a single researcher but rather from an iterative dialogue between three researchers. All the coding files have been kept and are available upon request.

3.3.11 Subgroup analysis and investigation of heterogeneity

We present pooled effect size estimates for the eight main outcomes for which we have data from at least two studies, namely yield, price, income from certified production, wages, total household income, wealth, illness and schooling. For outcomes for which we have only a single effect size, e.g. the number of training days of received by wage workers, we have incorporated these under the most fitting of the above eight outcomes.

3.3.12 Sensitivity analysis

Wherever the data allow it, that is for all outcomes with more the two effect sizes, we investigate the effect of stratifying the studies according to their risk of bias rating, their funding source and its peer-review status. We provide separate effect size estimates and appropriate confidence intervals, as well as full heterogeneity statistics for all of these groups for each outcome. By funding source we mean whether or not the study was funded, wholly or in part, by certification bodies or by NGOs engaged in helping others achieve certified status, i.e. by organisations who may have a vested interested in study results. Conversely, we consider studies to be independent if they are financed by public research bodies or government agencies. By peer-review status we mean whether or not a study was published in a peer-reviewed academic journal.

3.4 Deviations from protocol

Over the course of conducting the review we were not always able to implement everything what we had set out to do in our protocol (Oya et al, 2015). This section details the ways in which we have deviated from the protocol, and explains why these deviations occurred. All of these points have been mentioned before, but in the interest of transparency and convenience we gather them here again.

While had envisaged conducting hand searches in our protocol we had underestimated the amount of published material available on websites and non-standard databases. At the same time, the initial hand-searches identified a lot of potentially relevant material, which of course meant that the search process had to be conducted as widely as possible. Consequently, we spent many more person-hours on hand-searching that we had planned, leading not only to delays vis-à-vis our original timetable, but also constraining the resources available for all other stages of the review.

During the search process we did not search French-language databases. Given that other aspects of the search process were much more labour-intensive than expected, we did not have the resources to conduct full searches of social-scientific databases in French, German, Portuguese and Spanish. Our other searches were already turning up significant volumes of French-language material, which we screened for inclusion, meaning that French-language material was not excluded.

Given the importance of Latin America as a site for CS and Germany as a market for certified goods, we decided to prioritise German, Portuguese and Spanish databases. We therefore decided not to search other, specifically French-language sites, as we judged the probability of missing relevant material to be lowest there.

While our original scope had included organic certification, provided that socio-economic outcomes were reported on, we later dropped studies that reported only on organic certification. Our searches had turned up large amounts of material that dealt with socio-economic outcomes of organic certification, meaning we faced numbers of studies that far exceeded our limited resources. At the same time, the theory of change for organic certification is notably different to the other CS examined here. We consider the socio-economic effects of organic certification to be an important topic, richly deserving of its own review, especially because many socio-economic outcomes of organic certification are unintended. In the context of our review, however, the inclusion of organic certification was not only unmanageable in terms of the amount of additional material this entailed, but also would have diluted the focus of our research questions, and added another conceptual layer to the theory of change (i.e. expected links between different types of inputs and the production process). As a result we made the decision to drop all studies that deal only with organic certification and do not report on other CS. However, we retained studies that look at organic certification in conjunction with at least one other certification system. Many of the studies we did include analyse such double- or triple-certification.

We also had to conduct an additional, unplanned, round of coding for inclusion under RQ2. The number of studies we had included for RQ2 was far too large to allow for a meaningful synthesis of the evidence. As explained above, we therefore undertook another round of coding to eliminate all studies that did not provide substantive evidence on the three key themes we had identified as central barriers and facilitators.

Despite identifying such large amounts of material during the search process, the number of studies that were includable for RQ1 was smaller than we had initially expected at protocol stage. Our quantitative analysis was therefore faced with a lack of data, taking into consideration the potential large number of outcomes to analyse. At the same time, we also had less resources available than we had expected. This meant we had to carefully choose which outcomes to analyse for each study. For most studies this was not an issue as they only report on a small number of indicators. Some studies however, report on large numbers of variables. In these cases, as explained above, we chose to extract data for those variables for which we also had other studies available, so as to maximise the number of meta-analyses. We also took care to extract variables that measured constructs in similar ways to other studies, as the underlying similarity of measurement is a key assumption of meta-analytic methods. At the same time we focused on the most important indicators of key stages of the hypothesised causal chain. The scarcity of both data and resources meant we were not able to report impact estimates for all for all of the outcomes listed in the protocol and above.

This scarcity of data also affects the amount of moderator analysis we were able to undertake. In our protocol we had identified a number of different types of potential moderator variables to explore. Considering a wide range of moderators is only possible with a sufficient number of comparable studies reporting on the same outcomes though. With only a few studies available for each outcome, it is not especially meaningful to employ large numbers of moderators, as this leads to small cell sizes in each subgroup and thus instable findings. For the most part, we therefore confined ourselves to using the CS as a moderator, and to highlighting pattern of results with regards to different crops or regions as appropriate. Moderator analysis should not be confused with sensitivity analysis though. We have conducted thorough sensitivity analyses for all finding for which we have sufficient data.

4. Results

4.1 Description of studies

4.1.1 Results of the search

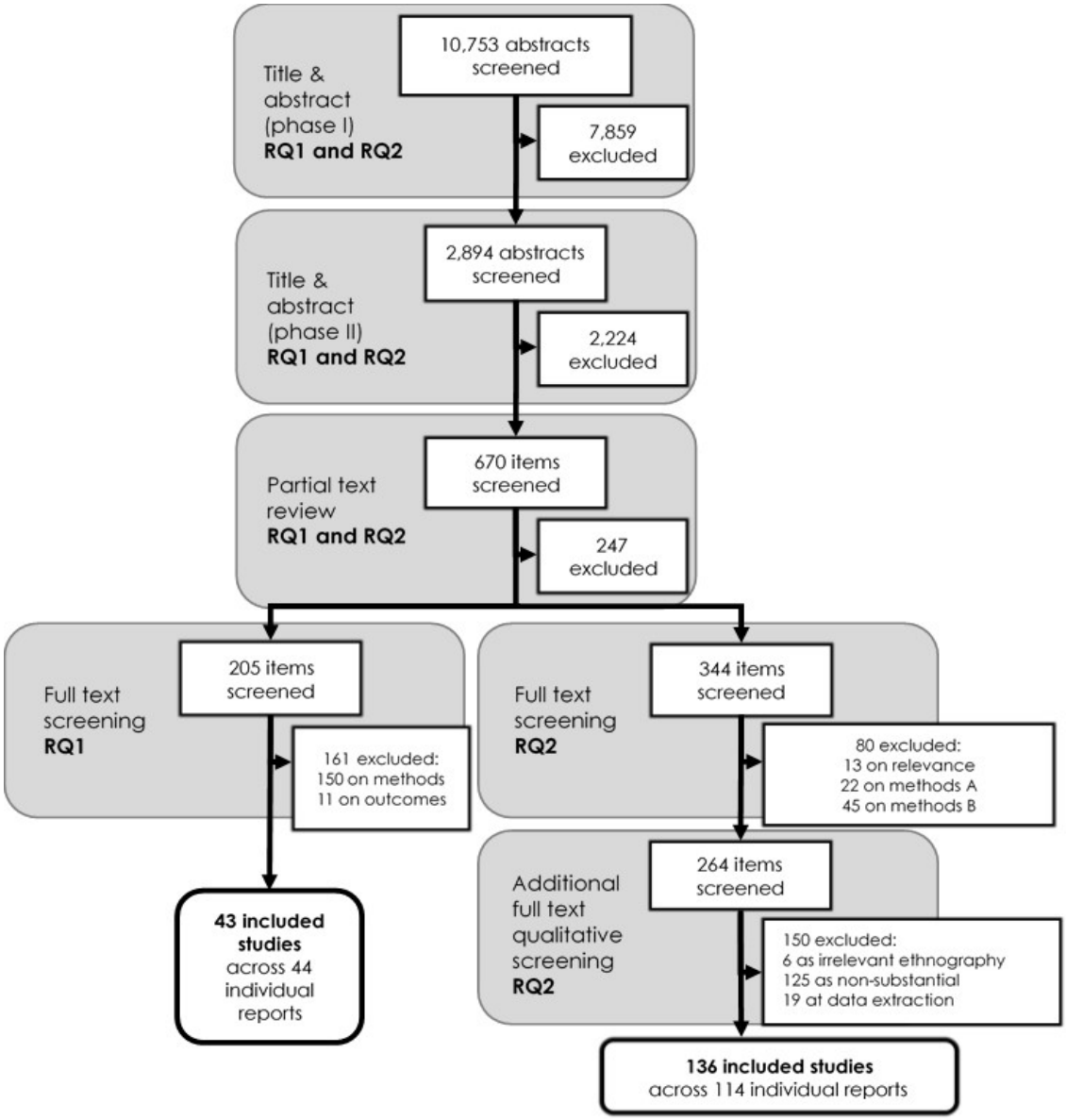
The initial search returned 10,753 papers. The approach used standardised search codes to retrieve papers from 15 databases; a further 31 databases and websites were manually searched, using a narrower selection of key search words for those with limited search functions. 10 German databases and websites were also searched.

This initial full list of papers was screened on abstracts by 4 research assistants, bringing the list down to 672. After full-text retrieval, an initial partial-text review was undertaken to exclude items not based on primary data, resulting in 247 papers for the full-text screening. At this point, papers with and without some sort of quantitative component were split evenly.

For the full-text screening, papers were assessed consecutively against separate criteria for review questions 1 and 2. At full text stage 205 papers were screened for review question 1, at which point 161 papers were excluded, while 344 papers were screened for review question 2 in two separate coding rounds with increasingly rigorous criteria, leading to the exclusion of a further 239 papers.

The resulting final list included 43 studies in 44 individual papers for review question 1 and 136 studies from 114 individual papers for review question 2; 20 studies were included for both research questions. **Figure 6** below demonstrates this process. The majority of our material comes from research reports, working papers, book chapters, and theses. Of the 43 studies included for RQ1 12 were published in journals (which we use as a proxy for peer review), while of the 136 studies included for RQ2 58 were taken from academic journals.

Figure 6: PRISMA diagram



4.1.2 Included studies

This review evaluated the effects of certification schemes for sustainable agricultural commodity production with a focus on socio-economic outcomes. Overall, included studies assessed 10 different certification schemes. Each of these schemes may have multiple standards as discussed in Section 1.

Table 2 summarises the core intervention components of the main CS featured in this review. As explained in Section 1, CS are best understood as bundles of intervention with varying emphasis on the various components. Even this distinction is somewhat crude, as the bundle of intervention that gets implemented depends on a host of contextual factors.

Table 2: Core components of included certification systems

Certification scheme	Components
4C (now Global Coffee Platform)	The Common Code for the Coffee Community (4C) includes criteria on price (price premium), labour, management, GAP, and market.
Better Cotton Initiative	Better Cotton Initiative's components include criteria on labour, management, and GAP.
Bird Friendly	Bird Friendly Coffee, a standard from the Smithsonian Migratory Bird Center, requires organic certification as a pre-requisite and otherwise focuses on environmental practices. Therefore its relevant components include price (price premium), GAP and labour criteria.
Ethical Trade Initiative (ETI)	The ETI standard is focused on labour and management components.
Fairtrade	Fairtrade includes criteria on price (guaranteed floor price), social premiums, labour, management, GAP, and market.
GlobalGAP	GlobalGAP includes components on labour and management.
Organic	Organic certification standards currently vary from country to country by regulatory agency. The standards are primarily focused on agricultural practices; relevant components include price (price premium), GAP and labour criteria.
Rainforest Alliance	The Rainforest Alliance standard includes components on GAP, labour, and management.
RSPO	The RSPO standard includes components on GAP, labour, management, and market.
Shop for Change	Shop for Change is an NGO-founded fair trade standard in India. Components include management, social premium ("capacity building premium"), labour, GAP and market.
Utz	The Utz standard includes components on GAP, labour, management, and price (price premium).

As seen in **Figure 7**, Fairtrade is the dominant scheme, evaluated by over half (52%) of all studies included. Many studies assessed multiple schemes, either through comparison or because target groups were certified in more than one programme. The most common combination was Fairtrade and organic, a result of the high number of producers certified in both schemes. Overall, 16 of the studies included for RQ1 and 58 of the studies included for RQ2 deal with more than one certification scheme. While this distribution does not represent the universe of CS and their relative significance in terms of volume of production certified or outreach across countries, these results clearly suggest there is a very substantial interest in Fairtrade more than in CS generally.

Figure 7: Certification schemes assessed by included studies[47]

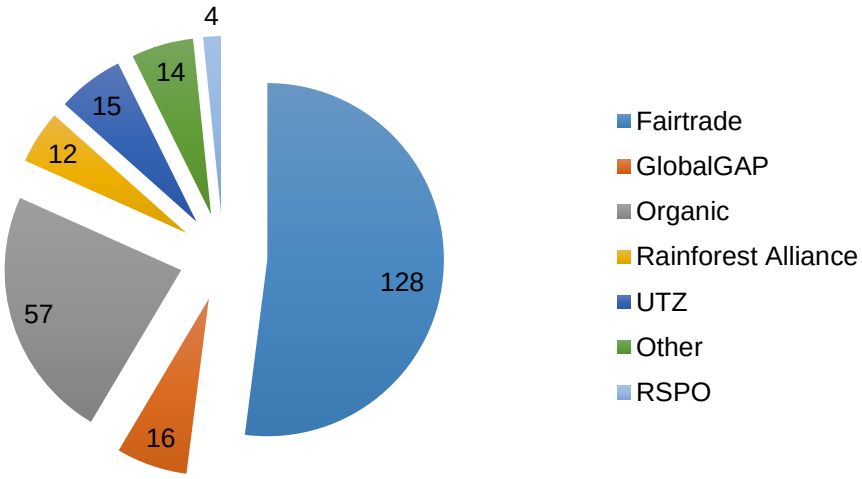
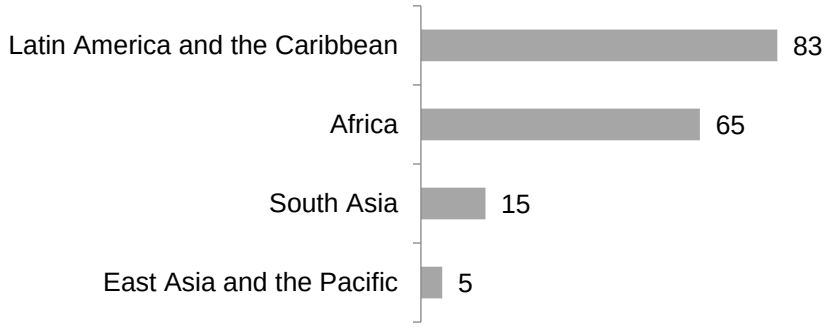


Figure 8: Regions of included studies[48]



[47] N=244. This figure includes studies included for both review questions. In this figure, N sums above the number of included studies because some studies assessed more than one CS.
[48] N=168. This figure includes studies included for both review questions. In this figure, N sums above the number of included studies because some studies took place in more than one region.

Figure 8 shows the geographical spread of studies, which took place in 30 countries, over three-quarters of which are in two global regions: Africa and Latin America and the Caribbean. This result is not surprising since a significant number of studies on Fairtrade certification focus on Latin America where this CS has a prominent presence. However, the particular focus on Latin America is at odds with the fact that the number of Fairtrade participants in Africa and the Middle East constitutes 64% compared to 20% in Latin America (data from Fairtrade 2016).

Coffee was the most commonly assessed commodity, evaluated by 38% of studies (**Figure 9**). This is consistent with the fact that coffee dominates Fairtrade (49% of all certified farmers and workers in 2015, according to Fairtrade, 2016), which is the most studied CS of all, and is also a key commodity for other CS like Utz. Studies included in the 'other' category assessed various nuts, grains, and honey.

Figure 9: Commodities evaluated by included studies[49]

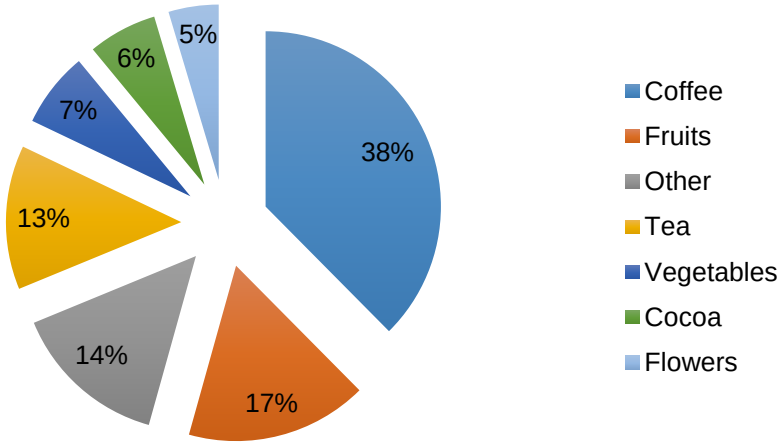


Figure 10 demonstrates the increase in publication of both qualitative and quantitative studies assessing certification schemes over the last decade. A particularly quick rise is seen in qualitative studies, whose overall amount dwarfs that of the publication of quantitative studies. The very rapid increase since 2010 for both quantitative evaluations and qualitative research reflects a combination of two trends. First, there has been an absolute increase in the research produced on CS, mirroring the growing interest in and expansion of standards for agricultural production. Second, a larger proportion of this research may be of higher quality and therefore able to meet the strict inclusion criteria adopted in this systematic review. This is welcome and shows that both the research community and organisations commissioning impact evaluations for CS are increasingly aware of the standard requirements for rigorous impact evaluation. This is especially clear in the plot of quantitative studies (RQ1) which shows that it has been only since 2008 that higher-quality study designs have emerged. Moreover, during the period 2014-2016 the number of quantitative studies for RQ1 doubled.

[49] N=171. This figure includes studies included for both review questions. In this figure, N sums above the number of included studies because some studies evaluated more than one commodity.

CS are associated with multiple interventions. Each CS may apply a set of standards for which multiple requirements are applied to participating producers. As discussed in Section 1, it is sometimes difficult to distinguish between CS in terms of their interventions because the overlaps are increasingly important as more CS converge towards applying a common pool of standards across different areas of social sustainability. Data were extracted from each study to determine which types of interventions were mentioned as linked to the CS under study.

Figure 10: Cumulative number of studies published per year[50]

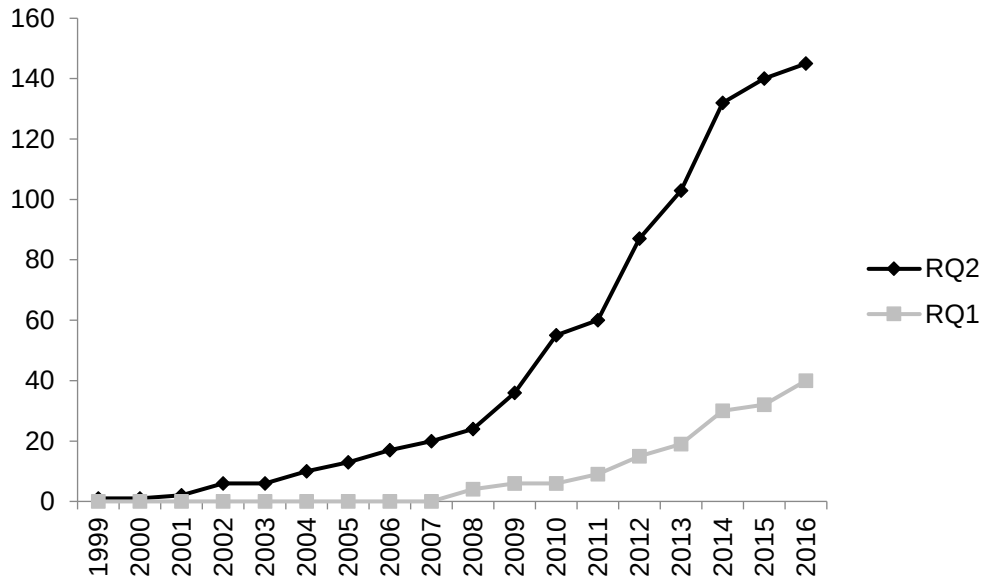
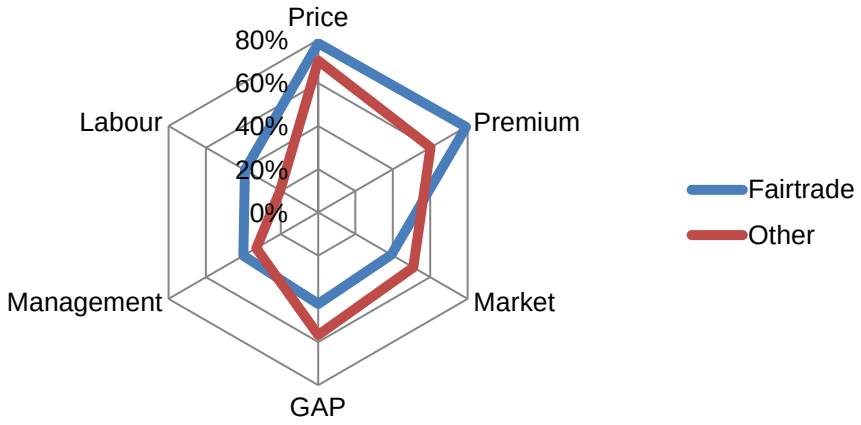


Figure 11 below shows a diamond for the proportion of studies reporting on Fairtrade (the most studied CS) and other CS that consider interventions across key types: labour standards, price, premium, market interventions, good agricultural practices (GAP, better farming), and management support (especially to POs). The diamond chart does corroborate the relative overlap among CS as they all tend to cover similar areas of intervention. However, Fairtrade stands out for its use of price and the premium as leading interventions in their engagement with small producers. Other CS place slightly more emphasis on market access (and branding) interventions, e.g. the act of certification which gives producers access to new markets that demand certain standards, or the creation of niche markets directly linked to the CS, as well as support to ensure traceability for labelling purposes and more lucrative markets.

[50] N=187. This figure includes studies included for both review questions. The sum is higher than total included because a) some studies were included for both review question 1 and review question 2 and b) some review question 2 studies were published across several papers (and therefore, on some occasions, several years).

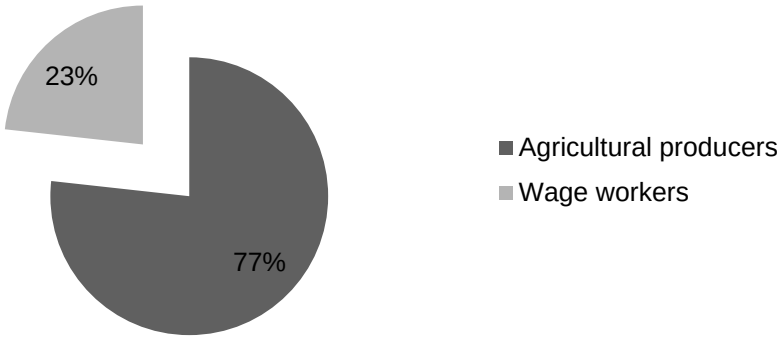
Figure 11: Incidence of types of interventions per CS[51]



Review question 1

CS target anyone directly involved in the production of agricultural commodities as well as other actors in the supply chain. This review is concerned with those participating in direct production, therefore both producers (farmers) and wage workers, whether employed by large-scale plantations or smallholder farms.[52] As seen in **Figure 12**, agricultural producers were the focus population in the majority of studies. This is despite the fact that most CS with social sustainability standards include labour standards. However, this pattern may reflect the fact that both the claims made by a number of CS and the researchers interested in certification for agricultural commodities tend to be focused on agricultural producers. The number of standards and associated requirements is also higher for farmers than for workers.

Figure 12: Study populations[53]



The descriptive summary for studies included for review question 1 is provided in Annex B. As the table there shows, the two most commonly used study designs

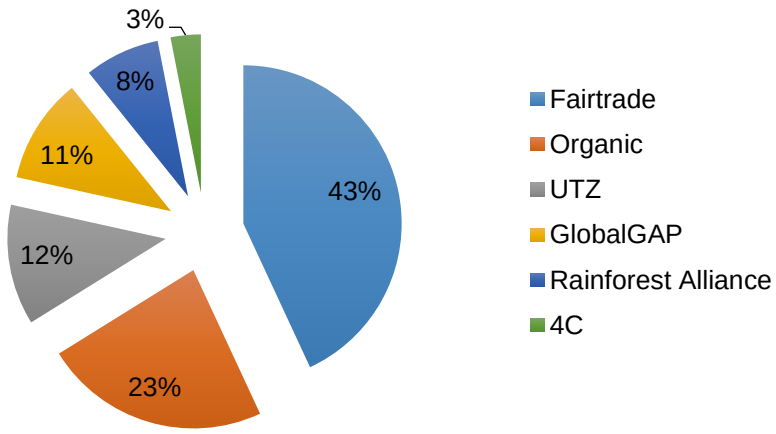
[51] N=159. This figure includes studies included for both review questions.
[52] The category wage workers includes all workers as reported on in the included studies. In theory, this includes paid family workers. In practice, however, the included studies focus on workers employed outside of their own household.
[53] N=43. This figure includes studies included for review question 1 (quantitative).

were controlled before-and-after and ex-post controlled observational studies, especially the latter by a large margin. Propensity score matching, difference-in-difference, and variations on least squares analyses were the most common methods of analysis.

We consider studies to have been independently financed if study authors make clear that the study was funded entirely without drawing on funds provided by either certification bodies or NGOs involved in providing certifications. Most commonly, independent research was financed either by governmental and intergovernmental agencies or by universities. We erred on the side of caution and have only labelled studies as independently financed when the source of finance is clear. Four studies (Becchetti & Gianfreda, 2008; Becchetti et al, 2008; Becchetti et al, 2011; Roy & Thorat, 2008) were labelled as not independently financed as no source of funding was acknowledged by the authors

Figure 13 and **Figure 14** below provide some insight into the range of certification schemes and countries in review question 1 studies. Fairtrade is still the dominant certification scheme of interest; combined with organic it represents two thirds of the included studies, a very large proportion, considering that there are multiple standards and CS present in LMICs. It is also noteworthy that the other CS represented here with a significant number of studies, i.e. Utz and RA, both have some strong similarities in terms of interventions, aims and ethos with the combination Fairtrade and organic. The majority of the studies are located in the global regions of Africa and Latin America and the Caribbean, with a small handful in South Asia.

Figure 13: Certification schemes assessed by studies included for RQ1[54]



[54] N=65. This figure includes studies included for review question 1 (quantitative). In this figure, N sums above the number of included studies because some studies evaluated more than one CS.

Figure 14: Regions of studies included for RQ1[55]

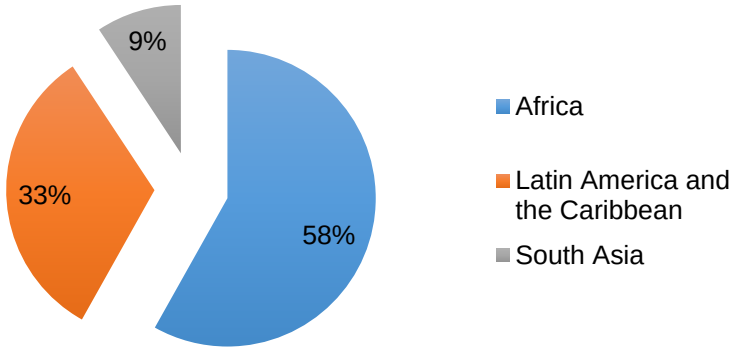


Figure 15 identifies the types of interventions reported on by included studies. The most common intervention was, by far, market interventions; 97% of studies included for RQ1 assessed a market intervention. Labour was the least evaluated type of intervention, with only 28% of studies clearly reporting on it. This is consistent with the lower proportion of studies focused on wage workers, compared to farmers. These figures should be interpreted with some caution due to the way this information is often reported. In many studies, it is not explicitly clear whether it is reporting on a certain intervention or not, nor whether the intervention, if reported on, is standards related (i.e. coop extension support provided by governments). Additionally, when a study evaluates more than one certification scheme, it may evaluate a type of intervention for one scheme but not the other(s).

Figure 15: Types of interventions assessed by included studies[56]

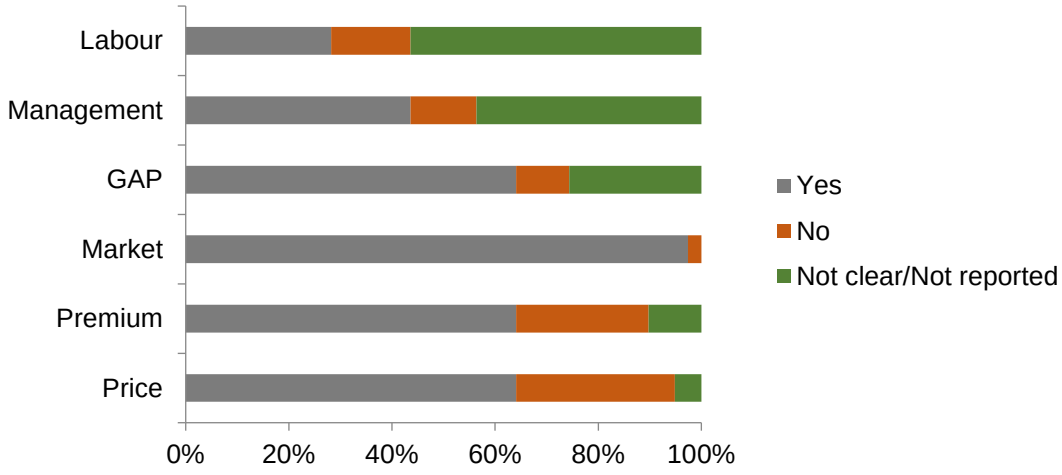
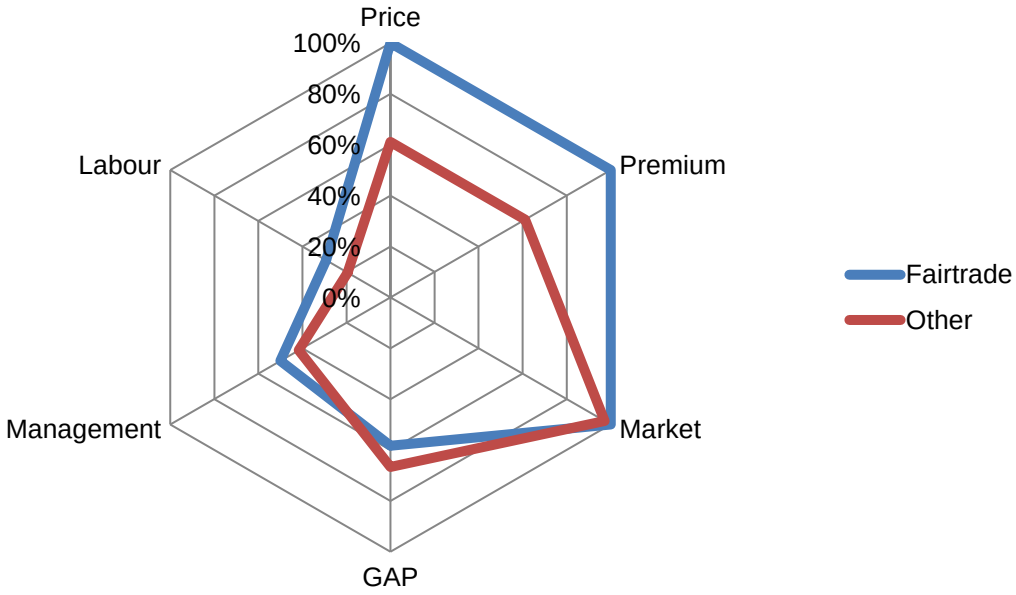


Figure 16 shows a diamond for the proportion of RQ1 studies reporting on Fairtrade (the most studied CS) and other CS that consider interventions across the main types. In this instance studies tend to report more on market interventions across all

[55] N=43. This figure includes studies included for review question 1 (quantitative).
[56] N=43. This figure includes studies included for review question 1 (quantitative).

CS, including Fairtrade. It is also notable the lack of focus on labour interventions, not surprising given the scarcity of RQ1 studies reporting on labour outcomes and standards.

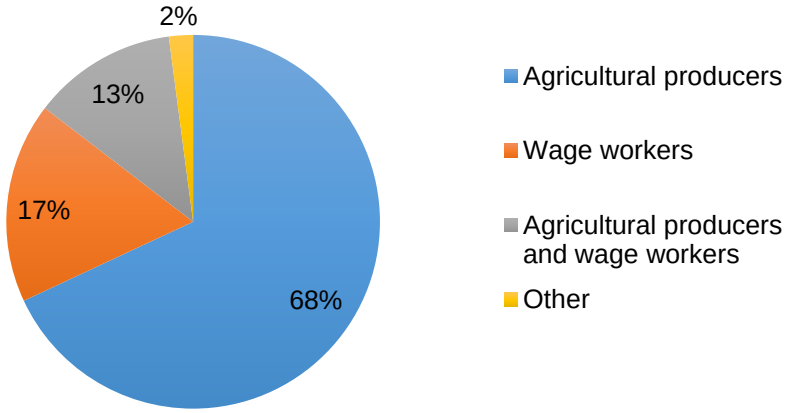
Figure 16: Incidence of types of interventions per CS for RQ1 studies[57]



Review question 2[58]

Table and **Table** summarise descriptive information for studies included for review question 2. A number of studies assessed multiple schemes and/or multiple commodities, and several also took place in more than one global region. Full details reported by study are found in Annex B.

Figure 17: Study populations[59]



[57] N=43. This figure includes studies included for review question 1 (quantitative).
[58] The tables and description in this section include studies that were included for review question 2 only as well as those included for both review question 1 and review question 2
[59] N=136. This figure includes studies included for review question 2 (qualitative).

As with studies included for RQ1, most studies evaluated agricultural producers (see **Figure 17)**. The “other” category here refers to several studies that assessed the impact of RSPO on relationships between large RSPO-certified companies and uncertified local landowners, specifically on their ability to assert land rights and access clean water.

Studies were identified in 4 global regions: South Asia, Africa, Latin America and the Caribbean, and East Asia and the Pacific. 12 certification schemes were assessed for 23 different agricultural commodities.

Table 3: Certification schemes by commodity and region[60]

Programme	Commodity (n=224)	Region (n=217)
Better Cotton Initiative	Cotton n=1	South Asia n=1
Bird friendly	Coffee n=3	Latin America and the Caribbean n=3
ETI	Flowers n=1 Vegetables n=1 Fruit n=1	Africa n=3
Fairtrade	Banana n=16 Cocoa n=8 Coffee n=49 Flowers n=7 Tea n=20 Other n=28	Africa n=40 Latin America and the Caribbean n=68 South Asia n=11
GlobalGAP (EurepG.A.P., PublicG.A.P.)	Banana n=1 Flowers n=1 Vegetables n=6 Fruit n=5 Others n=2	Africa n=6 East Asia and Pacific n=1 Latin America and the Caribbean n=4 South Asia n=1
Organic	Coffee n=35 Tea n=4 Banana n=3 Cocoa n=2 Other n=5	Africa n=7 Latin America and the Caribbean n=38 South Asia n=4
Rainforest Alliance	Cocoa n=2 Coffee n=4 Tea n=5	Africa n=5 Latin America and the Caribbean n=5 South Asia n=1
RSPO	Oil Palm n=4	East Asia and Pacific n=4
Shop for Change	Cotton n=1	South Asia n=1

[60] N for both columns sums higher than total included studies for review question 2 because many studies evaluated more than one commodity and/or took place in more than one region.

Programme	Commodity (n=224)	Region (n=217)
UTZ	Cocoa n=3 Coffee n=7 Tea n=1	Africa n=8 East Asia and Pacific n=1 Latin America and the Caribbean n=2
Various VSS	Coffee n=1 Flowers n=1 Vegetables n=2 Fruit n=1	Africa n=2 Latin America and the Caribbean n=1

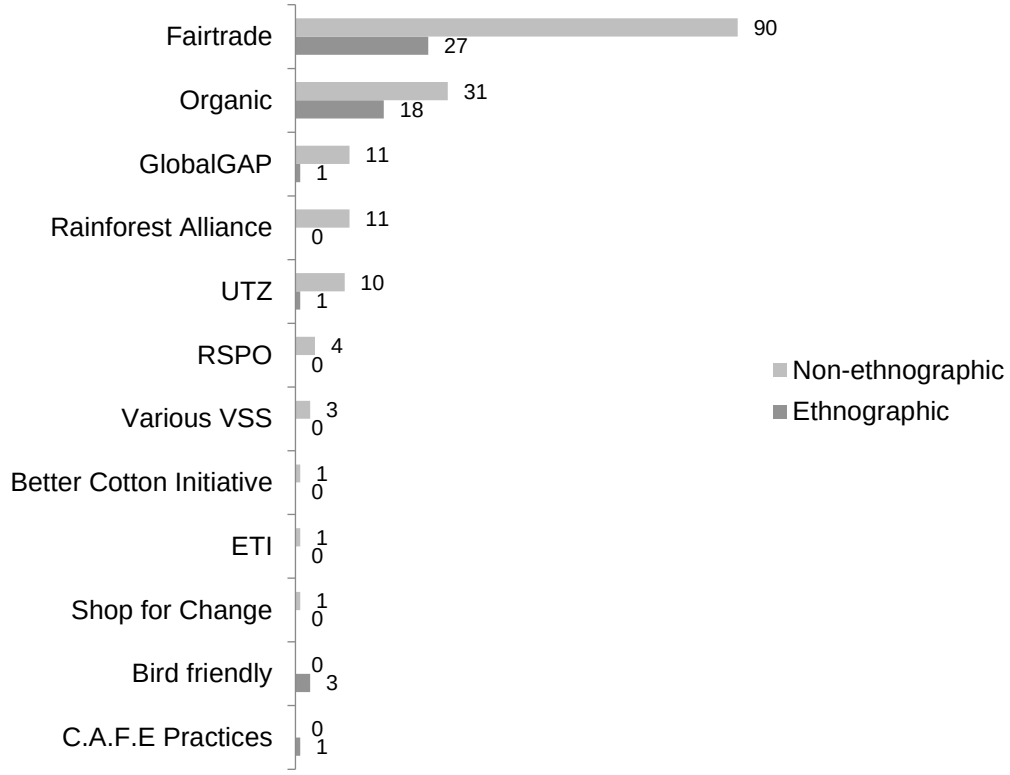
Evidence on all three major themes was identified for most certification schemes, though several (Better Cotton Initiative, RSPO, and Shop for Change) only had evidence on 2 out of 3 schemes. The most commonly reported on theme was implementation dynamics, quite close to other contextual factors and barriers and facilitators.

Table 4: Certification schemes by evidence theme[61]

Programme	Evidence on implementation dynamics (n=190)	Evidence on distributional dynamics (n=141)	Evidence on other contextual factors and barriers and facilitators (n=183)
Better Cotton Initiative	n=1	n=0	n=1
Bird friendly	n=3	n=1	n=1
ETI	n=1	n=2	n=3
Fairtrade	n=105	n=87	n=102
GlobalGAP (EurepGAP, PublicGAP)	n=8	n=5	n=11
Organic	n=47	n=31	n=44
Rainforest Alliance	n=10	n=8	n=6
RSPO	n=3	n=0	n=2
Shop for Change	n=1	n=0	n=1
UTZ	n=10	n=6	n=9
Various VSS	n=1	n=1	n=3

[61] N for each evidence theme sums higher than total included studies for review question 2 because some studies evaluated more than one CS.

Figure 18: Certification schemes by methodology[62]



The majority of the studies included for RQ2 utilised non-ethnographic research methods (**Figure 18**). Ethnographic research methods, when used, primarily focused on Fairtrade and organic certifications. This ratio is actually slightly higher than that at the full-text screening phase, in which 34 papers were ethnographic and 204 papers were non-ethnographic, suggesting a slightly higher proportion of high quality, relevant ethnographic studies.[63]

Figure 19 shows the types of interventions reported on by included studies[64]. The most common interventions are price and premium, while the fewest studies looked at market interventions. **Figure 20** shows a diamond for the proportion of RQ2 studies reporting on Fairtrade (the most studied CS) and other CS that consider interventions across the main types. In this instance studies tend to report much less market interventions across all CS, including Fairtrade. Price and premium interventions are then most often reported, with substantial overlap between Fairtrade and other CS. In this case labour interventions receive more attention than in RQ1 studies.

[62] N=214. The total sums higher than total included studies for review question 2 because some studies evaluated more than one CS.
[63] The graph shows more than 34 ethnographic studies due to the fact that some papers included more than one study.
[64] As discussed in section 1.3, this list of types of interventions is consistent with the taxonomy of four main groups used for the Theory of Change. This list allows to single out particularly common interventions such as price guarantee, good agricultural practices, and support to management of organisations and producers.

As with the graph for RQ1 studies, these figures should be interpreted with caution, however, due to the way this information is often reported. In many studies, it is not explicitly clear whether it is reporting on a certain intervention or not, nor whether the intervention, if reported on, is standards related (i.e. coop extension support provided by governments). Furthermore, it is not clear whether studies do not report out of neglect or a different focus, or because the intervention was not taking place to report on. If anything, these observations emphasise how messy evaluation of certification systems can be.

Figure 19: Types of interventions assessed by included studies[65]

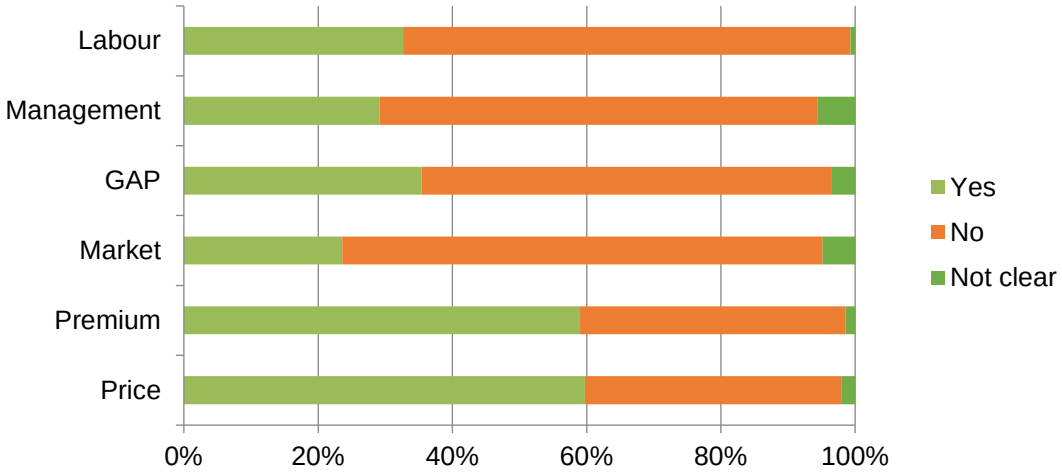
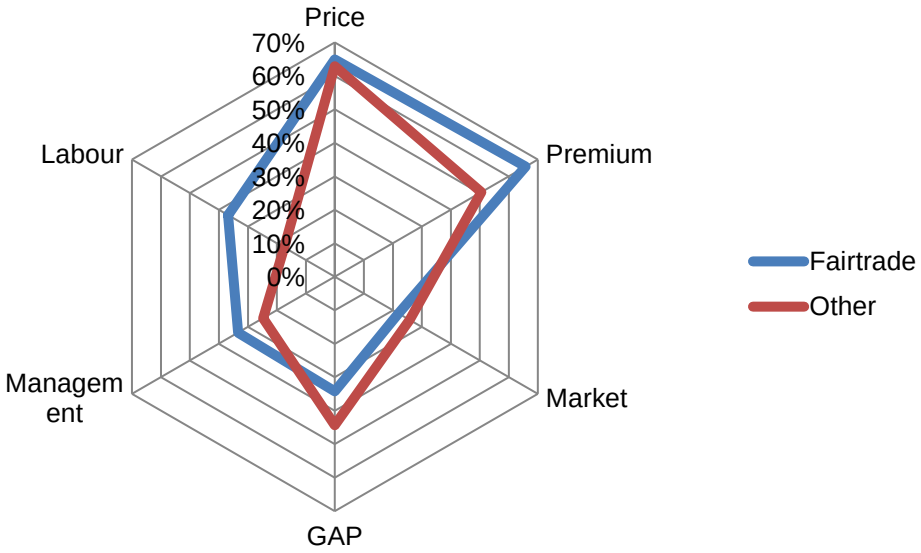


Figure 20: Incidence of types of interventions per CS for RQ2 studies[66]



[65] N=136.
[66] N=136.

4.1.3 Excluded studies

Review question 1

During the full-text screening, studies were excluded for RQ1 on the basis of evidence and methodological grounds. At that stage 11 studies were excluded for not assessing relevant outcomes, while 150 studies were excluded for not using an experimental or quasi-experimental design with adequate controls for confounding. In particular concerns about the adequacy of efforts to address selection bias and confounding led to the exclusion of many studies, though in some cases this was also down to weak reporting practices, see also Section 6.2. As discussed in the results section below, four studies were later excluded from the meta-analysis due to being rated as having critical risk of bias. High risk of bias studies were retained in the analysis.

Particular examples of prominent excluded studies include CEVAL (2012), which was prepared by a German research institute for two NGOs involved in certification. The study identified treatment and control groups over a variety of settings but is a prime example of what one might call ‘naïve control’ in that it makes no attempt whatsoever to control for selection bias to establish equivalence between groups. An example of a study that probably could have met inclusion criteria but had to be excluded for lack of reporting is Nelson and Martin (2013). The authors state that they used both PSM and DID to control for confounding and establish group equivalence, but provide no evidence as to whether and how these methods were applied, nor do they report other basic statistics. Another example of such a lack of reporting is COSA (2013) who state that they employed an IV model but then go on to report only mean differences between treatment and control rather than any regression outputs. Other studies by contrast failed to meet even the most the most basic of inclusion criteria. For instance, La Roche (2012) conducts an impact assessment without a control group.

Review question 2

During the full-text screenings, studies were excluded for RQ2 on methodological grounds; 45 studies inadequately reported on data collection and sampling, and 22 studies inadequately reported on a research question or objective. 157 studies were also excluded for not providing relevant substantive evidence on any of the three evidence themes.

4.2 Risk of bias in included studies

The assessment of the risk of bias is important for identifying and understanding issues in analysis and reporting that might have an impact on the results and conclusions of this review. As this report asks two research questions based on different methodologies, two separate risk of bias assessments were conducted for studies included for RQ1 and those included for RQ2.

4.2.1 Review question 1

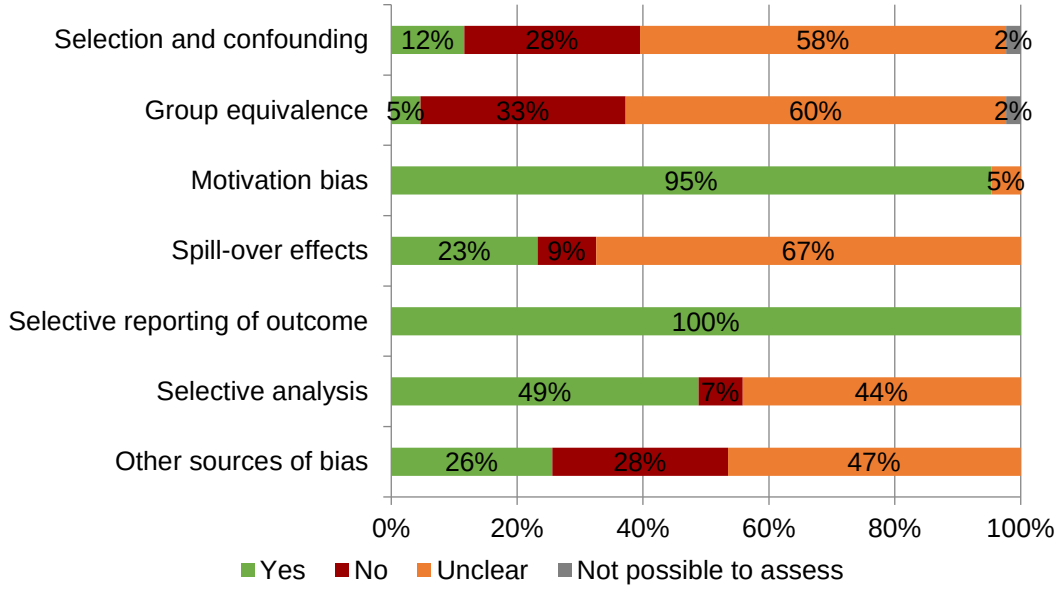
For studies included for RQ1, risk of bias was assessed through a tool built on the bias assessment tool developed by Waddington et al (2014), which categorises

studies as either low, medium or high risk of bias. Studies were assessed against the following seven elements:

1. Selection and confounding: Overall bias was very high here; only one study was judged to have adequately controlled for selection bias.
2. Group equivalence: Again, only one study had a fully adequate method of analysis to ensure comparability of treatment and control groups throughout the study and prevent confounding.
3. Motivation bias: All but two studies adequately protected against motivation bias caused by the process of being observed.
4. Spill-over effects: 29 studies were not adequately protected against performance bias, either through geographical and/or social separation.
5. Selective reporting of outcome: All studies were judged to be free from evidence of the selective reporting of outcomes.
6. Selective analysis: Only half (19) of the studies were judged to be free from the suggestion of biased exploratory research methods.
7. Other bias: Other forms of bias not captured in the categories above. A common example is doubt about measurement.

Figure 21 summarises the risk of bias assessment for all studies included for RQ1[67]. Please note that in the diagram a 'yes' means that the study fulfilled the criteria in this regard.

Figure 21: Summary of risk of bias across studies included for review question 1



The full risk of bias results for RQ1 studies can be found in Annex C. As explained in Section 3.3.3 above, the risk of bias assessment for Cramer et al (2014) presented a potential conflict of interest and was therefore handled solely by an external consultant.

[67] N=43.

4.2.2 Review question 2

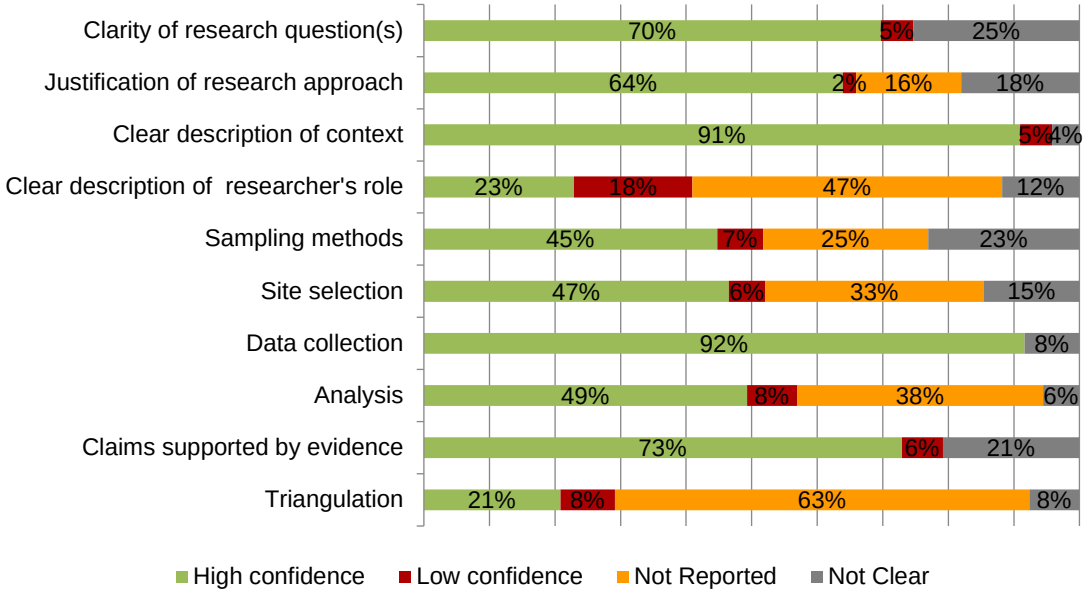
To perform the critical appraisal for qualitative studies included for RQ2, this review used an adapted CASP (2006) tool alongside the approach used in Waddington et al (2014), which combined both substantive and reporting-based criteria. Included studies were assessed on the following elements:

1. Clarity of research question*:* Only 7 studies did not report on research questions; a quarter (36) did not report on them clearly; generally these studies provided some information on goals but did not go so far as to provide clear research questions.
2. Justification of research approach: 92 studies satisfactorily justified their research approach; most others either failed to report on it or did not do so clearly. Only 3 studies were assessed as failing to provide justification.
3. Clear description of context: All of studies reported on research context, almost all of which (131) did so clearly and appropriately.
4. Clear description of researcher's role: Only 33 studies adequately reported on the researcher's role in the study.
5. Sampling methods: Less than half of the included studies (64) reported clearly on their methods for sampling.
6. Site selection: As with sampling methods, less than half (67) reported clearly and separately on how site were selected for research.
7. Data collection: All studies reported on data collection methodology, and only 12 did not report them sufficiently.
8. Analysis: Just under half of the studies (71) adequately reported method of analysis.
9. Claims supported by evidence: While most studies adequately supported their claims through evidence, 9 studies failed to provide a sufficient link between the data and conclusions.
10. Triangulation: While many studies utilised multiple methods of data collection, only 30 studies directly reported on the use of triangulation.

A summary of the results of the critical appraisal for RQ2 studies is presented in **Figure 22**. The results demonstrate a wide heterogeneity in the confidence we can place in these studies; this result is likely at least partially due to the inherently difficult nature of fairly and comprehensively assessing study and reporting quality across a wide range of qualitative studies. The screening process had ensured sufficient information on context as studies without any substantive evidence on key themes and with very weak reporting were excluded at full-text stage. However, on other criteria a high proportion of studies did not report anything relevant. The critical appraisal also depends on the nature of the publications. Ethnographies are likely to have better confidence ratings because there is a tradition of reporting more on methods and the research process, including issues of reflexivity and triangulation, than in more rapid qualitative studies that are conducted with a focus on gathering perceptions from participants or understanding the certification process and main contextual issues. Many such studies report limited information on methods, usually in the form of a list of methods of data collection applied with not much information on justification for selection of research sites, let alone for the sampling of

respondents interviewed through focus groups, semi-structured interviews or participatory techniques. Overall RQ2 studies do especially well in terms of describing the contexts, especially the location and characteristics of participants and their engagement with certification. They also generally report enough on the data collection tools and make research questions explicit. Reflexivity (researcher's role), triangulation, sampling methods, data analysis methods and justification for site selection tend to be unreported or simply not considered at all. Again this is partly due to the nature of the reports/papers, which tend not to be long ethnographies and are generally focused on presenting findings.

Figure 22: Summary of critical appraisal across studies included for review question 2[68]



A synthesised quality assessment was not undertaken for RQ2 due to the wide range of methodological approaches involved, including ethnographic studies. Instead, a full quality assessment for each study can be found in Annex C.

4.3 Assessment of publication bias for RQ1

The results of a meta-analysis depend, of course, on the individual studies that entered into the analysis. An analysis that systematically excludes certain studies or types of studies will be biased. In systematic reviews the particular concern is that the search process might have excluded certain classes of publications (for instance, non-peer reviewed publications) or that the search may have failed to find studies that were not published in academic journals. Compounding the potential problem, academic journals themselves have a tendency to publish studies that report statistically significant findings, as well as to report the findings of large and well-funded studies. Publication bias, and the related availability bias, in systematic

[68] N=136.

reviews therefore typically arises when reviews have failed to find studies that were either not published in standard academic channels (e.g. research reports, working papers and theses) or not published at all, and when these studies differ systematically from included studies. The typical assumption of this model is that omitted studies will have effect sizes that are substantively smaller and/or less likely to be statistically significant than included studies, in which case the pooled effect would be inflated (for a detailed discussion of these issues, see Borenstein et al, 2009).

In short, these biases can result from an incomplete search process, where non-standard publication types are omitted. As outlined in the previous chapter, in this review we therefore searched not only general and subject-specific scientific databases, but also extensively searched sources of grey literature such as databases of theses and the websites of certification bodies, research institutes and NGOs working on related issues. Through our advisory group we also reached out for unpublished material held by experts and certification bodies. Our included studies for RQ1 come from working papers and research reports, as well as from academic journals, and the range of material included for RQ2 is wider still. A comprehensive search process is the best protection against publication bias, and there is therefore little *a priori* reason to expect that publication bias affects our findings.

Nonetheless, it is worthwhile to investigate the possibility of bias more formally. We do this in three ways. First, wherever we have sufficient data we conduct sensitivity analysis on our findings to see whether the results from studies that were published in peer-reviewed academic journals are different from those published through other channels.[69] The results of these sensitivity analyses are reported with the synthesis results for each outcome in Section 4.4. Second, as part of our risk of bias evaluation, we assessed whether studies included for RQ1 had likely engaged in selective reporting of outcomes, which can bias results as unreported outcomes are likely to be substantively small and statistically significant. We found few indications that included studies had selectively reported their findings. Only three studies included for RQ1 we rated as problematic as having probably engaged in selective reporting of finding (see Annex C). Third, we use funnel plots to map our findings and attempt to visually identify possible instances of bias. Below we present funnel plots for all outcomes for which we have at least eight data points, i.e. for income from certified production, wages and total household income.[70] The funnel plots below (see **Figures 23**, **24**, and **25**) graph the estimated effect size (i.e. the SMD) for each included study (on the x-axis) against its associated standard error (on the y-axis). The standard error, which is in large parts determined by sample size, is used here as the measure of precision with which each study estimates the effect size. Larger

[69] We use journal publication as a shorthand for peer-review, but the correlation is of course not perfect. Not all journals use peer review, and other publication types, such as research reports or book chapter, may well undergo a process of peer review prior to publication.
[70] Funnel plots are increasingly difficult to interpret as the number of studies falls and we do not find it sensible to product such plots for less than eight studies (see also Sterne et al. 2011). The plots presented were created using the *metafunnel* command in Stata 14.2.

and more precise studies will be shown towards the top of the graph, where the standard error tends towards zero, while smaller and less precise studies are displayed towards the bottom of the graph. A vertical black line shows the estimated pooled effect, while dashed lines show the edges of the confidence interval within which we would expect the majority of estimates effects to fall. Please note that the scale differs from graph to graph for presentational reasons.

An asymmetric distribution of studies in a funnel plot can indicate the presence of bias. However, interpreting funnel is a subjective activity and is only made more ambiguous when dealing with small numbers of studies, as we do here. Moreover, publication bias is one of the possible causes of asymmetry in a funnel plot. Observed heterogeneity and asymmetry may for instance be due to real, substantive differences across studies and interventions (Sterne et al, 2011). Asymmetry in this case may also be due to the exacting inclusion criteria used, which meant that less reliable evidence was excluded. Of the plots below, only the funnel plot for wages shows a striking asymmetry.

Figure 23: Funnel plot for income from certified production

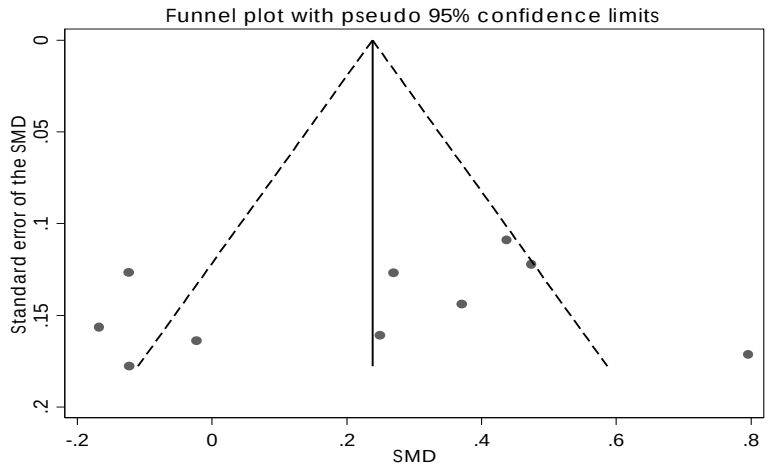


Figure 24: Funnel plot for household income

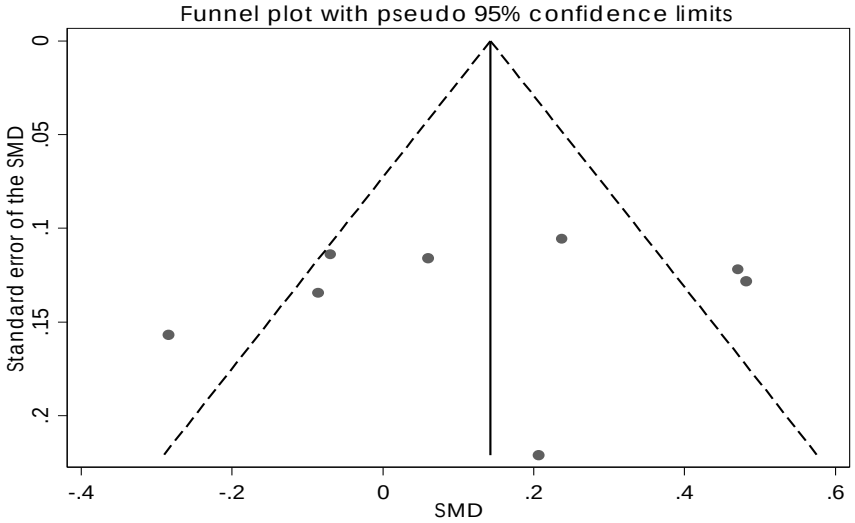
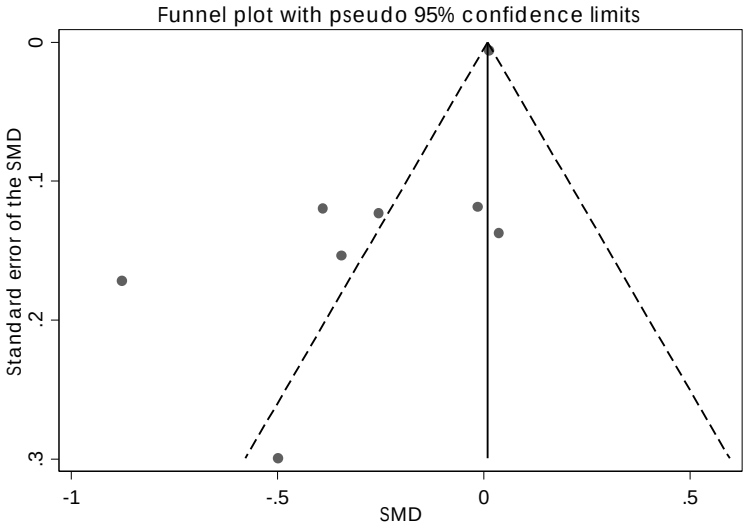


Figure 25: Funnel plot for wages



As the analysis of funnel plots is open to misinterpretation when only few studies are present, we also performed Egger's tests for all outcomes for which we produced funnel plots. The tests were performed using the *metabias* command in Stata 14.2. A significant test result allows us to reject the null hypothesis of no small study effects and thus may be indicative of publication bias. As **Table 5** shows, only the test for wages returns a significant result (defined as a p-value for bias that is smaller than or equal to 0.05), indicating the possibility of publication bias. However, we do not believe this result to be indicative of publication bias in this case. The skew present in the funnel plot is towards negative results. For either publication bias or the file drawer problem (i.e. the underreporting of certain results) to be present publications would have to favour reporting negative and null results over positive ones, and researchers would have to systematically favour reporting negative over positive results – both of which are usually considered implausible assumptions, and especially so given the literature we are dealing with (see also Dickersin, 2006).

Table 5: Results of Egger's test for small study effects

Outcome		Number of studies	Co-efficient	SE	t-stat	p-value	95%-CI lower bound	95%-CI upper bound
Income from certified production	Slope	10	0.70	0.61	1.15	0.282	-0.70	2.10
	Bias	10	-3.32	4.31	-0.77	0.464	-13.26	6.62
Wages	Slope	8	0.03	0.01	2.10	0.080	-0.004	0.05
	Bias	8	-2.21	0.72	-3.07	0.022	-3.97	-0.45
Total household income	Slope	8	0.35	0.53	0.67	0.527	-0.93	1.64
	Bias	8	-1.67	4.10	-0.41	0.697	-11.70	8.35

4.4 Quantitative synthesis of results (RQ1)

4.4.1 Overview

This section presents the results of the meta-analysis we performed to address RQ1. Results are presented in form of forest plots. All plots are labelled with scales and notes on the direction of effect for ease of interpretation, with zero meaning no effect. Please note that for presentational reasons the scales differ across forest plots. For each study, the plots give the extracted effect size, along with bars to represent the 95%-confidence interval, while a diamond at the bottom of the diagram gives the calculated overall effect and its 95%-confidence interval. Exact figures for both are given on the right of the diagram, and we provide a significance test for each of the calculated overall effect sizes.[71] The effect sizes are sample size-corrected SMDs (i.e. Hedge's g), meaning that effects are expressed in standard deviation units. Boxes around the point estimates for each study represent the weight of each study in the calculation of the overall effect size. For each forest plot we present the appropriate diagnostic statistics for heterogeneity, namely the estimate of between studies variance T^2 (expressed in the same metric as the effect sizes themselves, albeit squared) and the I^2 statistic, which describes the share of observed variance that is due to real differences in effects sizes between studies. For completeness, we also report the heterogeneity Chi-squared and associated p-value (i.e. the test for significance of the Q statistic), although these may have limited practical value (Borenstein et al, 2009). In interpreting effects, readers are reminded that all studies included in this section have study designs that are intended to come as close to a causal attribution of effects to certification as possible.

For each outcome we have combined effect sizes from studies that measure similar underlying constructs, but do so in a wide variety of different geographic and socio-economic contexts, and frequently using a range of different study designs and analytical methods. Our syntheses combine studies across different crops, each with their own agronomic peculiarities, political economy of production and supply chain structure. We also combine studies looking at different CS and CS types, even though these operate according to quite distinct theories of change. It is therefore not surprising that we encounter high levels of heterogeneity in our analyses.

We explore this heterogeneity through the use of moderator and sensitivity analysis. Typically, moderator analysis will employ extrinsic, methodological, and substantive moderators (Lipsey, 2009). However, the limited amount of data we have available means that we are constrained in the number and types of moderators we can meaningfully employ. Moderator analyses conducted on limited numbers of studies must be interpreted with caution as the number of studies in each category will be small. We therefore consider it prudent to limit the moderator analysis we undertake to the most important difference between studies. As we consider the type of certification scheme to be the most important difference in terms of the aim of this

[71] Effects are judged to be statistically significant if a test whether they are different from zero returns a p-value smaller than or equal to 0.05. Please note that we also report a p-value for the test of significance of the heterogeneity Chi-squared statistic. These should not be confused.

review, we have generally limited ourselves to providing an analysis that separates out studies by the certification schemes they examine. We provide such an analysis whenever we have at least five studies for a given outcome.[72] For the outcome for which we have the most data – income from certified production – we also present the results separated by crop type. Wherever appropriate in terms of data availability we also comment on differences across studies due to the crop or region examined for other outcomes.

Wherever data allow, we conduct sensitivity analysis to examine whether results are sensitive to the inclusion of studies rated as having a high risk of bias, whether studies that were fully independently financed provide different results to studies that were at least partly financed by certification bodies or NGOs involved in certification, and whether results differ for studies published in peer-reviewed academic journals as opposed to other forms of publication that lack a peer-review mechanism.[73] To be clear, studies published in peer-reviewed publications are not necessarily of higher quality than studies published in other ways. We do not regard peer review in and of itself as a quality criteria – this is what risk of bias assessments is for. We do not present formal sensitivity analyses for either assets/wealth or illness as we only have two studies available for each of these outcomes.

We have synthesised effects along the causal chain, moving from intermediate to endpoint outcomes. We begin with yields, i.e. the amount of crop produced per unit of land. If certification changes how much producers have available to sell in the market this can help raise (or reduce) their incomes. Yields alone of course cannot determine income, however, as revenue is a function of quantity and price. We therefore next look at the effects of certification on price. Prices are of course also an input for many certification schemes, but here we understand prices as an outcome, in that we look at the prices actually received by producers. Both better prices and better yields can individually or in combination lead to higher incomes. We then turn next to incomes from certified production. This refers to the incomes received by certified farmers compared to non-certified farmers for the sale of certified produce. In other words, it disregards any other (non-certified) produce sold by the same farm as well as all other income generating activities. However, farmers are not the only group of agricultural producers who could potentially benefit from certification. Our next outcome is therefore the impact of certification on the wages received by workers employed in certified production. Regrettably, this is the only data we have on wage labourers and all other outcomes once more deal with certified farmers.

Turning thus once again to certified farmers, income is a key outcome, as just discussed. However, what matters most to households is not simply what happens to income from certified production, but whether overall household income rises. Total household income has a complicated relationship to income from certified production and a rise in the latter does not necessarily mean a rise in the former as well. From

[72] An exception is schooling, where the results of separating studies by certification scheme can be read from the main forest plot, thus making a separate forest plot unnecessary.
[73] We regard studies as fully independently financed if they were funded by a public research funding institution or an international organisation.

household incomes we turn to measures of wealth, before moving on to illness – for each of which we have only limited data. Our final outcome deals with the education of children living in households engaged in certified production.

Overall, we have extracted 53 separate effect sizes from 29 different studies. We have conducted meta-analyses for each outcome for which we have at least two effect sizes. In total 44 effect sizes were used for eight main meta-analyses. The other effect sizes are mostly instances where we have only a single effect size estimate for a given outcome. These are referred to under the most appropriate heading. Other effect sizes were excluded from the analysis to preserve the independence of effect sizes, as explained in Section 3 above. As discussed in the last section, 43 studies met the inclusion criteria for RQ1. However, we were unable to extract effect sizes from seven studies (García et al, 2014; Kuit et al, 2016 on Uganda; Kuit et al, 2016 on Vietnam; Roy and Thorat, 2008; Ruben and Zuniga-Arias, 2011; van Rijn, 2016 on the Dominican Republic; van Rijn, 2016 on the Colombia). In all cases this was because studies simply reported too little information to allow us to calculate standardised effect sizes. No study was removed from the review just because we were unable to calculate an effect size. Rather results from such studies are referred to wherever relevant, i.e. under the appropriate headings below, but these studies did not contribute to the meta-analyses and thus to the pooled effect sizes. A further five studies, namely Barham and Weber (2012), Cepeda et al (2013), Nelson et al (2013), Rijn et al (2016, Ghana) and Stathers and Gathuthi (2013), were rated has having critical risk of bias and results from these studies will thus not be used in the analyses below, as their inclusion could distort results.[74] Lastly, as discussed in Section 3.3, we removed two studies from the meta-analysis to preserve the independence of effect sizes. Both Kamau et al (2010) and Ruben et al (2014) share part of their data with van Rijsbergen et al (2016) and thus cannot be considered independent. Their findings enter into the meta-analysis through the inclusion of van Rijsbergen et al (2016), which had the strongest research design and lowest risk of bias of the three studies.

For almost all studies we extracted all the data we could. Due to resource constraints however, this was not possible where studies reported on large numbers of variables (e.g Jena et al, 2012). In this limited number of cases we chose to focus on variables for which we had at least one other study (to allow for meta-analysis) and choose variables that were measured in similar ways to other studies, so as to minimise heterogeneity due to measurement issues, or came closer conceptually to the underlying construct (for instance we always chose to extract information on net rather than gross incomes, where both where available). Most studies with too many variables to allow for full extraction did not however end up contributing to the meta-analysis, as they either lacked the necessary information to allow for the calculation of effect sizes (e.g. Ruben and Zuniga-Arias, 2011) or we dropped to preserve the independence of effect sizes (Ruben et al, 2014). We have reported information on all outcomes for which we have extracted data and all extracted data has been

[74] Two other studies by Rijn et al (2016) are included in the analysis, but the Ghana study lacked a non-treated counterfactual and was thus rated as having critical risk of bias.

reported.

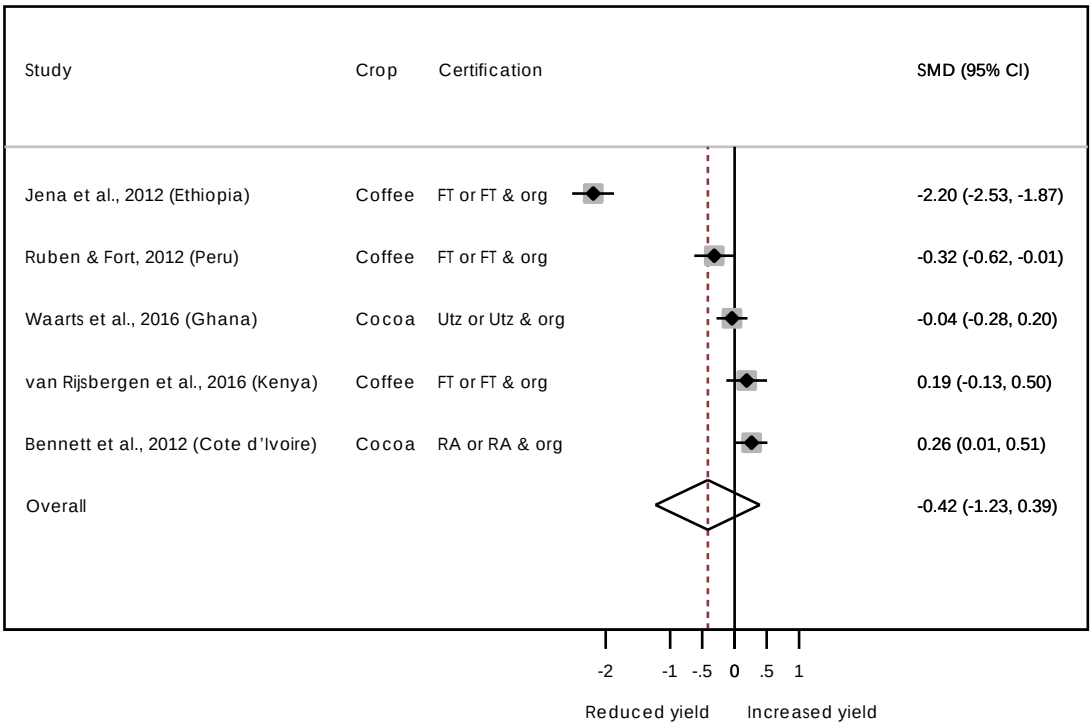
The effect size estimates used for the meta-analyses presented below have not been adjusted for unit of analysis issues. As outlined in 3.3.5 this is because the unit of analysis adjustments introduce additional assumptions and therefore uncertainty. We have instead presented unit of analysis-adjusted results in Annex F. The effects of adjustment are minimal, leaving the direction and statistical significance of pooled effects unchanged in almost all cases. The sole exception to this is the pooled effect for schooling, which remains positive but is no longer statistically significant once the adjustments have been applied.

Overall, the scarcity of data limits the confidence we can have in the results presented for RQ1. A small number of studies always runs the risk that findings might be idiosyncratic – a risk that even strict inclusion criteria, as well as moderator and sensitivity analyses, cannot fully eliminate. However, the results we find are in line with key elements of the qualitative synthesis, and an integrated synthesis of results is presented in Section 4.6 below. Also, as shown in **Figure 10** above more and more high-quality impact assessments are becoming available and future reviews will hopefully be able to draw on a wider base of reliable quantitative evidence.

4.4.2 Yield

In agricultural production yield – the amount of produce grown on a particular piece of land over a given period – is a central determinant of producers' incomes. We extracted data on yields from five included studies, as presented in **Figure 26** below. Across these five effect sizes we found a reduction in yields for certified farmers (SMD -0.42, 95% CI from -1.23 to 0.39), although the effect is not statistically significant (p=0.312). There is very substantial heterogeneity (I^2=97.5%). In part this is due to the presence of a clear outlier, Jena et al (2012), whose point estimate (SMD -2.2, 95% CI from -2.53 to -1.87) lies far to the left all other estimates. The other statistically significant results are Ruben and Fort (2012), who find only a modestly negative result, and Bennett at al (2012). Re-running the analysis excluding Jena et al (2012) produces a smaller pooled effect (SMD 0.03), with a 95%-confidence interval ranging from -0.22 to 0.27, meaning that the effect is not statistically significant (p=0.819). Of course, removing the outlier greatly reduces the heterogeneity of the result, with I^2 falling to 68.3%.

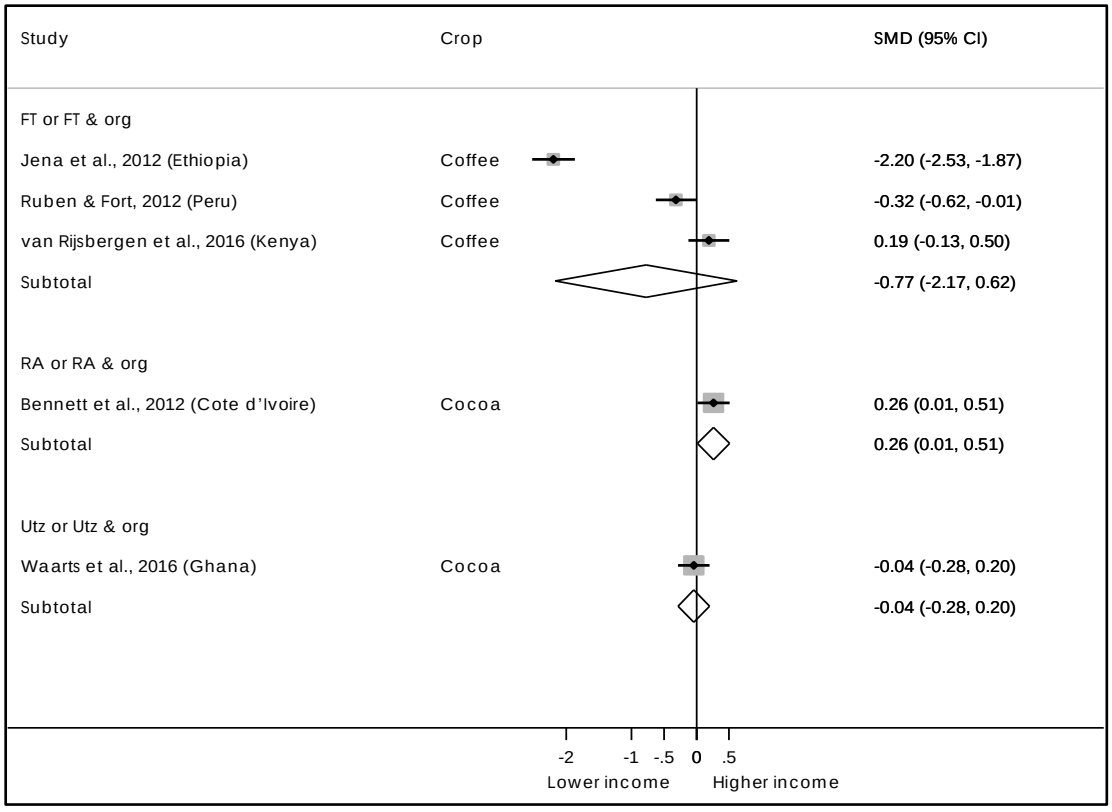
Figure 26: Forest plot for yield



Diagnostics: Heterogeneity chi-squared = 163.05 (d.f. = 4) p = 0.000. I-squared (variation in ES attributable to heterogeneity) = 97.5%. Estimate of between-study variance Tau-squared = 0.8309. Test of ES=0: z=1.01, p = 0.312.

The mixed picture extends also to studies for which we could not calculate effect sizes. Cepeda et al (2013) report significantly higher yields amongst Ecuadorian cocoa farmers, as do García et al (2014) for Colombian coffee growers, while Kuit et al (2016) find no effect on yield for Ugandan coffee farmers. Changes in yields can have a large variety of causes, from weather to the effects of tree age in coffee. Moreover, effects may be inconsistent across time. Remaining with the example of coffee, the 'stumping' or cutting back of old coffee trees will reduce yields in the short term but is vital to maintaining future yields once trees reach a certain age. Commenting on the concrete reasons for changes in yields requires a detailed examination of the farming system and contextual environmental variables in each instance. This information is not consistently reported across all included studies. We can thus only reflect the yield changes as reported in included studies and must remain agnostic as to the underlying causes in each case.

Figure 27: Forest plot for yield by certification scheme



Based on the data we were able to meta-analyse, we conclude that we have no evidence that certification significantly the raises the yields achieved by certified producers. CS differ in how much emphasis they place on agricultural practices. Utz and RA, for instance, emphasise training in 'good agricultural practices', which includes measures to raise and maintain the productivity of the land. Yet even these CS, along with many others, perhaps place more emphasis on quality improvements, with the expectation that this will lead to higher prices.

Table 6: Heterogeneity statistics for moderator plot

Category	Q	T^2	I^2	p-value for Q	# of studies
FT or FT & org	117.9	1.5013	98.3%	0.000	3
RA or RA & org	0	0	0	-	1
Utz or Utz & org	0	0	0	-	1

In **Figure 27** we use the certification scheme as a moderator and show results stratified by scheme. While most results are clustered around zero, the heterogeneity of effects for Fairtrade is striking. However as in the main meta-analysis above, this is mostly driven by the large negative effect provided by Jena et al (2012). Moreover, for both RA and Utz we only have one study each, which of course precludes any judgement about the variance in effect sizes for these schemes. Heterogeneity statistics for **Figure 27** are given in **Table 6**.

Table 7: Sensitivity analysis for yield

Category	SMD	CI lower bound	CI upper bound	Q	T^2	I^2	p-value for Q	# of studies
All studies	-0.42	-.23	0.39	163.05	0.83	97.5%	0.312	5
Risk of bias								
High	-1.117	0.998	-3.231	108.52	2.31	99.1%	0.00	2
Moderate	-0.019	-0.585	0.547	8.19	0.14	87.8%	0.004	2
Low	0.188	-0.127	0.503	0	0	0	-	1
Study independence								
Fully independent	-2.198	-2.525	-1.872	0	0	0	-	1
Not fully independent	0.029	-0.216	0.273	9.46	0.024	68.3%	0.024	4
Peer-reviewed publication								
Peer-reviewed journal	-0.775	-2.173	0.624	117.19	1.501	98.3%	0.00	3
Other publication	0.109	-0.187	0.406	2.94	0.0302	65.9%	0.087	2

Table 7 presents sensitivity analyses of studies reporting on yields. Stratifying studies by risk of bias does not affect results. It is worth noting, though, that Jena et al (2012) displays a high risk of bias. On the other hand, Jena et al (2012) is also the only study to be fully independently financed. Combining the non-independent studies yields a non-significant result. Studies that were published in peer-reviewed journals show an overall negative impact (SMD -0.78, 95% CI from -2.17 to 0.63), while studies that did not undergo peer review find a positive impact. However, the effect estimates for neither the peer-reviewed nor the non-peer-reviewed group are statistically significant. Moreover, as emphasised above, peer review is not by itself a guarantor of study quality.

4.4.3 Price

Many, if not most, CS do not primarily aim to increase the yields of certified producers. All however build centrally on the idea that certified production gives unique access to niche markets, namely those for certified produce. The assumption is that prices paid in certified markets should be higher than those attainable outside of such markets, as this is the prime motivation for undergoing the effort, and shouldering the costs, of certification in the first place. The effect on prices is a gross effect, that is, it does not take into account issues such as input costs or the costs of

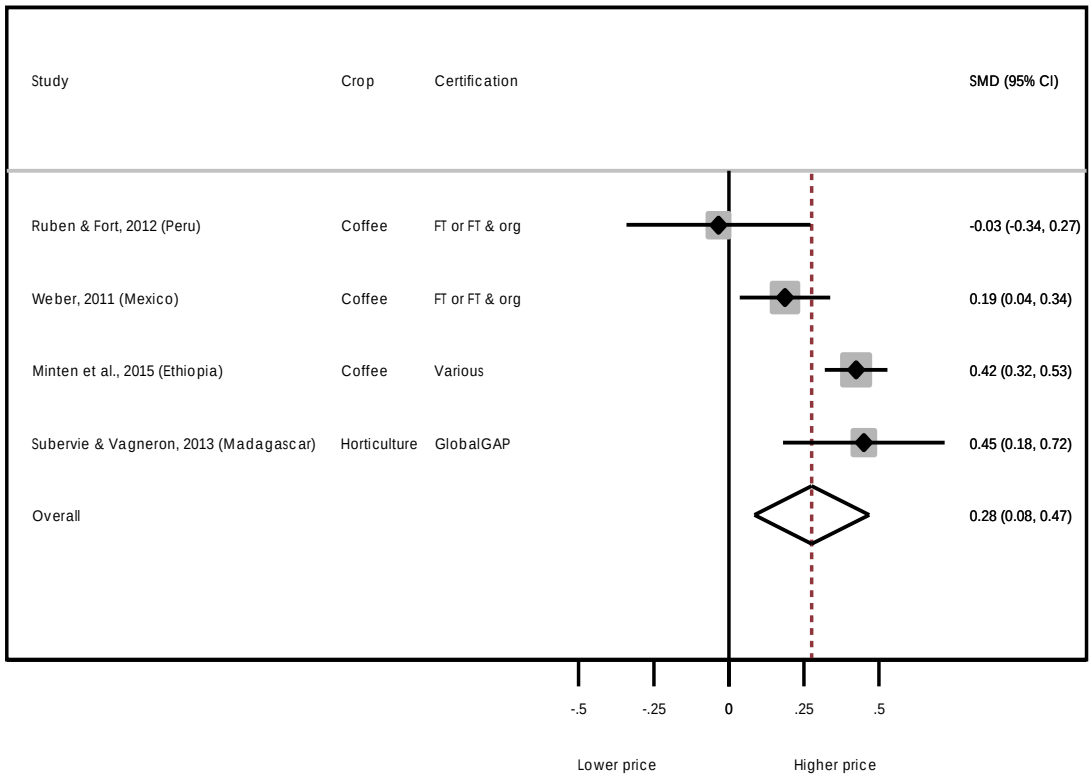
credit.[75] We look at both gross and net income from certified production, which takes all of these changes into account as far as possible, in the next sub-section.

Often higher prices are related to quality, and Balineau (2012) finds that certified cotton producers in Mali do produce higher quality cotton compared to non-certified producers. Some CS, prominently Fairtrade, go further than this and provide guaranteed minimum prices to 'insure' certified producers against some of the downside risks they face in selling their produce.

Only four studies provided information on price, as presented in **Figure 28**. The overall effect is an increase in the price (SMD 0.28, 95% CI from 0.06 to 0.47) and the effect is statistically significant (p=0.005). While there is less heterogeneity than for yields, it is still substantial (I^2=76.5%). There are however no clear outliers and the three positive effects are all statistically significant while a null result (Ruben and Fort, 2012) completes the picture. Based on the limited evidence available, certified producers are able to sell their produce for significantly higher prices than non-certified producers. This result is given further credence by one of the most positive effects found, Minten et al (2015), who cross-referenced data from a large-N survey with data on export sales taken from administrative sources, meaning that their results build on particularly large set of observations. On the other hand three of the four results are based on just one commodity, coffee, thus possibly limiting the external validity of these findings across other commodity types. Amongst the studies for which effect sizes could not be calculated the evidence is mixed. García et al (2014) and Ruben and Zúñiga-Arias (2011) find no significant effects on prices as a result of certification, while Weber (2011) finds that Mexican coffee farmers receive significantly higher prices. Moreover, studies do not always differentiate between the prices received for all produce sold and the prices received specifically for certified produce. This is because demand constraints mean that producers can typically only sell part of their output through certified marketing channels even though all of the production might be certified.
One effect size from Anteneh et al (2014), which looked at Fairtrade certified coffee farmers in Ethiopia, was not used in the analysis as the study did not have an uncertified control group, but rather compared the effects of double and triple certification to single certification. For Anteneh et al (2014) we calculated a statistically significant positive effect (SMD 0.24, with a 95%-confidence interval stretching from 0.09 to 0.39) for the addition of another certification to an already certified farmer. As we have so little evidence on price we do not conduct a moderator analysis for this outcome.

[75] For instance, credit costs may rise due to the need to finance higher levels of inputs, or the costs on inputs may fall due to a change in agricultural practice as a result of certification. Moreover, certification may result in access to cheaper credit.

Figure 28: Forest plot for price



Diagnostics: Heterogeneity chi-squared = 12.78 (d.f. = 3) p = 0.005; I-squared (variation in ES attributable to heterogeneity) = 76.5%; Estimate of between-study variance Tau-squared = 0.0271; Test of ES=0: z= 2.83, p = 0.005.

Table 8: Sensitivity analysis for price

Category	SMD	CI lower bound	CI upper bound	Q	T^2	I^2	p-value for Q	# of studies
All studies	0.28	0.08	0.47	12.78	0.0271	76.5%	0.005	4
Risk of bias								
High	0.450	0.181	0.719	0	0	0	-	1
Moderate	0.225	-0.012	0.462	11.97	0.0345	83.3%	0.003	3
Study independence								
Fully independent	0.345	0.171	0.520	6.92	0.0162	71.1%	0.031	3
Not fully independent	-0.034	-0.340	0.272	0	0	0	-	1
Peer-reviewed publication								
Peer-reviewed journal	0.207	-0.021	0.436	5.57	0.0260	64.1%	0.062	3
Other publication	0.424	0.320	0.528	0	0	0	-	1

Sensitivity analysis however shows these findings to be unstable, as demonstrated in **Table 8**. Removing the one study with high risk of bias, Subervie and Vagneron (2013), and looking only at studies with moderate risk of bias produces a pooled effect that is (just) statistically insignificant. Similarly, removing the one study that was not published in a peer-reviewed journal, Minten et al (2015), gives a pooled effect whose confidence interval just catches zero, making the results statistically not significant. On the other hand focusing only on studies that were fully independently financed, i.e. disregarding Ruben and Fort (2012), maintains the significant positive effect. Moreover, care must be taken in interpreting sensitivity analyses conducted with such a small number of effect sizes. For instance, as mentioned above, Minten et al (2015) were rated as moderate risk of bias and the study was published through a widely-respected international research institute (the International Food Policy Research Institute, IFPRI), so that the fact that it was not published in a peer-reviewed academic journal is not very meaningful in this case.

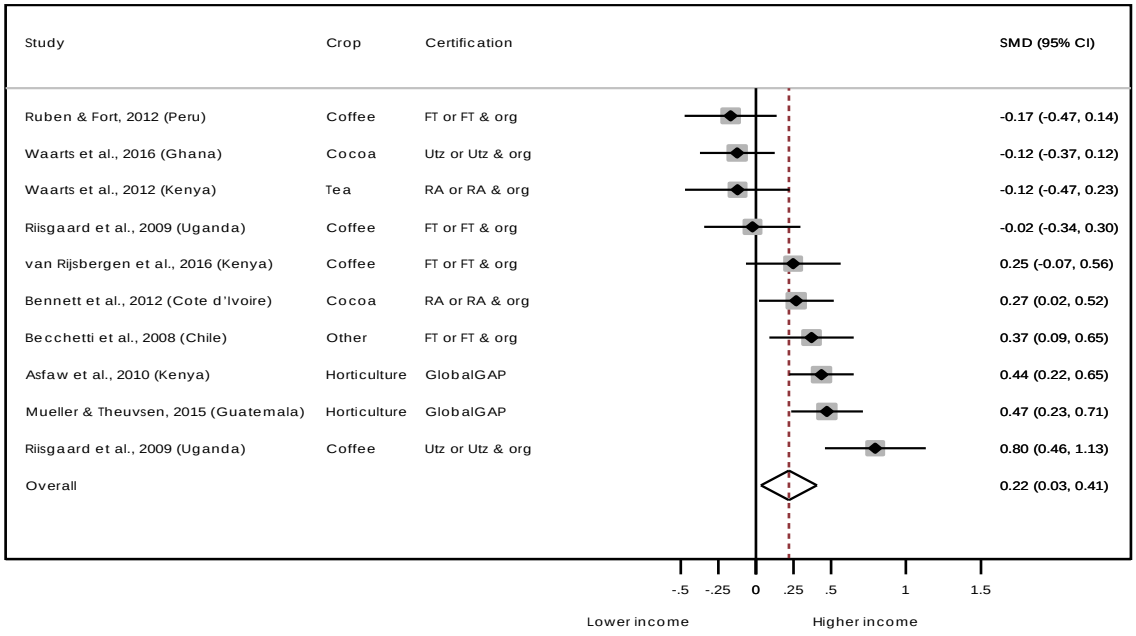
4.4.4 Income from certified production (net and gross)

While, as noted, higher prices are indeed a key measure of success for many CS, they are a highly imperfect one. The price they receive matters less to producers than the overall income they receive from engaging in certified production. The price paid per unit of certified output may not be a very meaningful indicator of increased wellbeing if only a small part of the certified output can be sold through certified

channels. We therefore now turn to the effect of certification on the income producers actually receive from their participation in certified production. The studies included here compare the income producers receive from the production and sale of a particular certified commodity, such as coffee, with the income received by otherwise equivalent groups producing the same commodity but lacking certification. It is important to note however that for many producers income from certified production will be only one source of income for the household. In Section 4.4.6 we also look at the effect on total household income. Similarly, not all certified produce can always be sold as such. Due to demand limitations in the markets for certified products, producers may only be able to sell part of their production as certified, despite the fact that all of it is certified. This situation is typical of coffee production for instance.

As shown in **Figure 29** we synthesised the effects from 10 studies on the income producers receive from engaging in certified production. The overall effect from certification in the evidence analysed is an increase in income from production of that commodity (SMD 0.22, 95% CI from 0.03 to 0.41). The overall effect is statistically significant (p=0.021). The results show substantial heterogeneity (I^2=77.6%), but there are no clear outliers. Bennett et al (2012, Cote d'Ivoire) and Ruben and Fort (2012) also provided effect size estimates for gross income from certified production but we chose to include the net estimates given by both studies instead. Similarly, Waarts et al (2016) also provided an estimated effect for profits from certified production per hectare, but we included net revenue from certified production instead, as this is more comparable to the included effects sizes.

Figure 29: Forest plot for income from certified production



Diagnostics: Heterogeneity chi-squared = 40.15 (d.f. = 9) p = 0.000; I-squared (variation in ES attributable to heterogeneity) = 77.6%; Estimate of between-study variance Tau-squared = 0.0692; Test of ES=0: z= 2.31, p = 0.021.

We combine studies reporting net incomes from certified production, with ones producing gross incomes. Arguably net incomes from certified production are a more meaningful indicator, as these take into account any additional costs that may have arisen as a result of certification. At the same time, net incomes are much more difficult to measure and are likely to be subject to greater measurement error. Removing the three studies reporting gross income, namely Becchetti et al (2008), Mueller and Theuvsen (2015) and van Rijsbergen et al (2016), still produces a positive pooled effect (SMD 0.154, with a 95%- confidence interval from -0.099 to 0.408). However the effect is therefore no longer statistically significant (p=0.233). As was to be expected, the heterogeneity of findings has also increased with I^2 rising to 82.4%. Amongst the studies that we could not extract effect sizes from, Cepeda et al (2013), García et al (2014) and Roy and Thorat (2008) all report higher incomes form certified production, looking at Ecuadorian cocoa farmers, Colombian coffee farmers and Indian grape growers, respectively. Kuit et al (2016) find no statistically significant effect on income for Ugandan coffee farmers and a reduction in income for certified farmers in Vietnam. Similarly, Ruben and Zúñiga-Arias (2011) find no statistically significant effect on incomes for coffee farmers in Nicaragua.

Figure 30: Forest plot for income from cert. prod. by certification

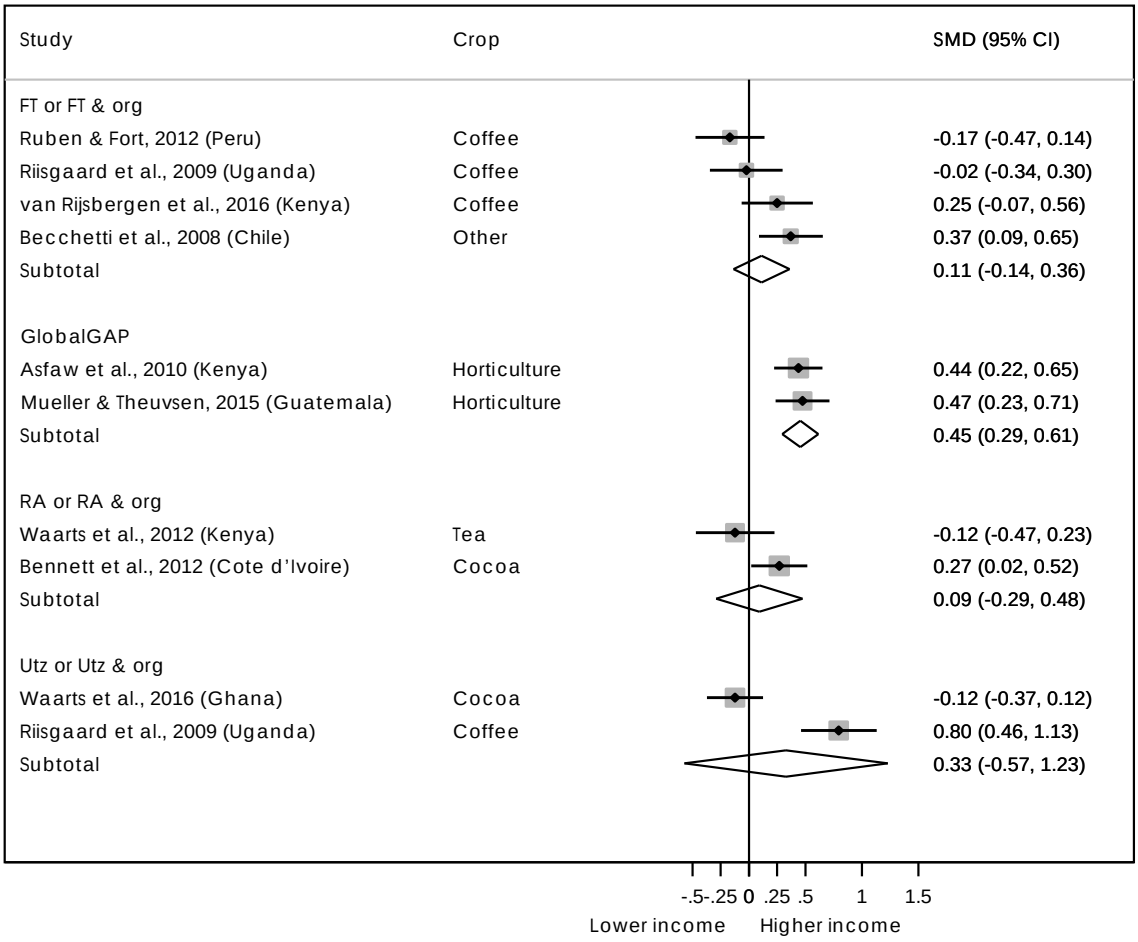
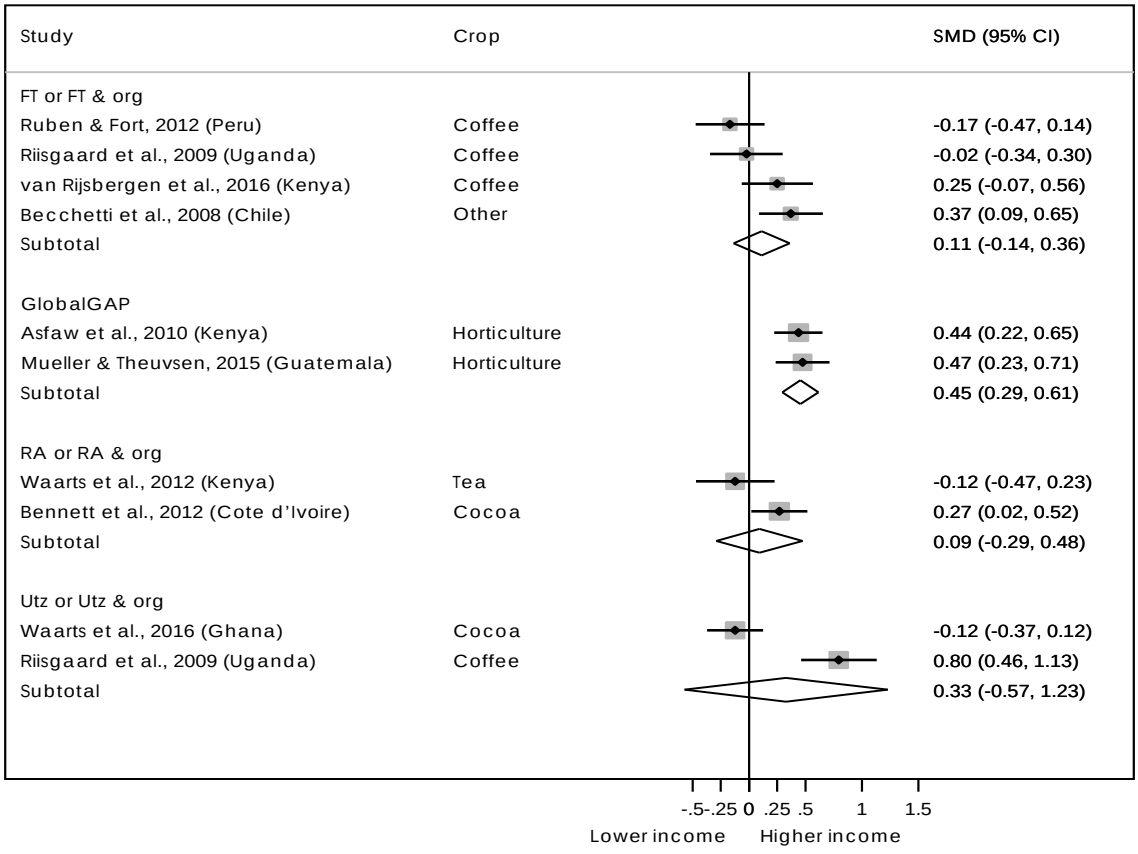


Figure 30 stratified the studies according to certification scheme. A clear positive and statistically significant impact emerges for GlobalGAP, where two studies give an SMD of 0.45, with a 95%-confidence interval ranging from 0.29 to 0.61. None of the other schemes show a statistically significant effect. Another noticeable feature is that for all schemes apart from GlobalGAP there are studies reporting negative effects, even though none of these are statistically significant. In the case of Utz, where the heterogeneity among studies is especially stark, this may be driven by differences between commodities, but Fairtrade shows substantial variation even though all three of the studies look at coffee. We report heterogeneity statistics for this forest plot in **Table 9**.

Table 9: Heterogeneity statistics for moderator plot (CS)

Category	Q	T^2	I^2	p-value for Q	# of studies
FT or FT & org	7.83	0.0392	61.7%	0.050	4
GlobalGAP	0.05	0	0	0.824	2
RA or RA & org	3.22	0.0531	69.0%	0.073	2
Utz or Utz & org	18.63	0.4008	94.6%	0.00	2

Figure 31: Forest plot for income from certified production by crop



To explore these issues further **Figure 31** presents the same information stratified by crop type.[76] Once again, the GlobalGAP certified horticultural producers see a statistically significant positive impact, as does the one study reporting on honey producers, Becchetti et al (2008). No clear pattern emerges for the other crops and none of the effects size estimates for the other crops are statistically significant, though the wide dispersion of effect size estimates is particularly noticeable in coffee. As above, heterogeneity statistics for the forest plot are given in **Table 10** below.

Table 10: Heterogeneity statistics for moderator plot (crop)

Category	Q	T^2	I^2	p-value for Q	# of studies
Coffee	19.57	0.1468	84.7%	0.00	4
Horticulture	0.05	0	0	0.824	2
Other	0	0	0	-	1
Cocoa	4.81	0.0613	79.2%	0.028	2
Tea	0	0	0	-	1

Table 11: Sensitivity analysis for income from certified production

Category	SMD	CI lower bound	CI upper bound	Q	T^2	I^2	p-value for Q	# of studies
All studies	0.22	0.03	0.41	40.15	0.0692	77.6%	0.00	10
Risk of bias								
High	0.151	-0.158	0.459	15.96	0.0795	81.2%	0.001	4
Moderate	0.206	-0.077	0.489	28.80	0.1027	82.6%	0.000	5
Low	0.249	-0.066	0.565	0	0	0	-	1
Study independence								
Fully independent	0.290	-0.025	0.606	8.74	0.0592	77.1%	0.013	3
Not fully independent	0.190	-0.045	0.426	27.58	0.0786	78.2%	0.000	7
Peer-reviewed publication								
Peer-reviewed journal	0.039	-0.370	0.448	3.46	0.0619	71.1%	0.063	2
Other publication	0.263	0.054	0.471	32.80	0.0702	78.7%	0.000	8

[76] The 'other' category refers to honey in this case.

Overall, the findings for income from certified production do not prove to be very stable (see **Table 11)**. Stratifying studies according to high (Waarts et al, 2012; Waarts et al, 2016; Becchetti et al, 2008; Asfaw et al, 2010), moderate (Ruben and Fort, 2012; Riisgaard et al, 2009; Riisgaard et al, 2009, Bennett et al, 2012; Mueller and Theuvsen, 2015) and low risk of bias (van Rijsbergen et al, 2016) produces positive but statistically insignificant effects for all three groups. Similarly, studies that were fully independently financed (Waarts et al,2012; Asfaw et al, 2010; Mueller and Theuvsen, 2015) and those that were not both show positive but statistically non-significant effects. The only difference emerges when looking at where studies were published. Separating out the two studies published in peer-reviewed journals (Ruben and Fort, 2012; van Rijsbergen et al, 2016) produces a positive and statistically significant effect for the remaining studies, while the pooled effect for the peer-reviewed group is no longer statistically significant.

4.4.5 Wages

Up to now we have discussed outcomes for independent agricultural producers. However this represents just one side of the story. The other group of direct producers who gain income from producing agricultural commodities are of course wage workers. Certified products are produced both by smallholder farmers and by large plantations, both of whom rely on wage labour, albeit to different extents. Smallholder farmers are not a homogenous group, but rather encompass a variety of different scales of production, and larger smallholders especially frequently hire wage workers in addition to family labour and other non-market interactions (Kevane, 1994; Oya and Pontara, 2015). Plantation agriculture is typically completely reliant on wage labour. Many certification schemes incorporate minimum requirements regarding the welfare of wage workers, though in some cases, such as GlobalGAP, this means little more than a requirement to conform to national labour laws.

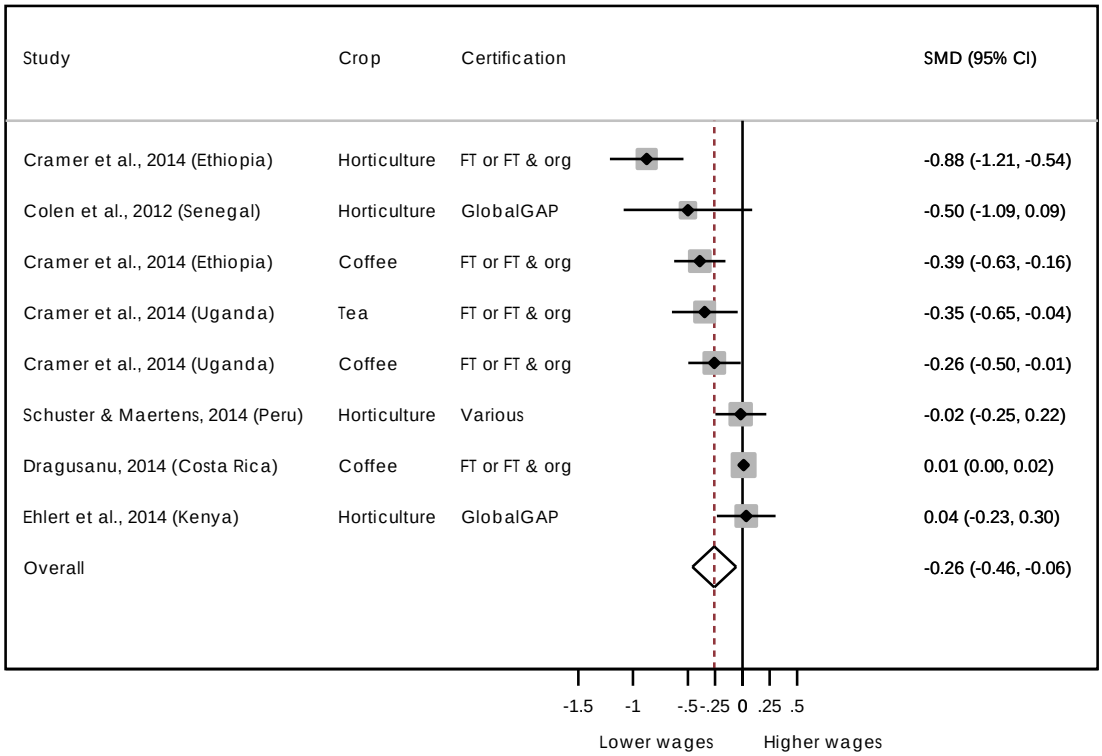
There is a longstanding concern about the effect of the business scale of the production unit on wages and non-wage working conditions (see for instance Cramer et al, 2014 and Ehlert et al, 2014). Given the scarcity of data we cannot address this concern here. We include studies that deal with the employees of large-scale production units and those that include workers employed on smaller farms.[77] Concretely, Colen et al (2012) and Schuster and Maertens (2014) look at the employees of larger companies, while Cramer et al (2014) and Dragusanu (2014) examine both workers on smallholder farms and on plantations.[78] Ehlert et al (2014) focuses on workers on small farms. However, Ehlert et al define small farms as farms employing 15 workers or less, which would not fall within many peoples' understanding of the term 'smallholder'.

As **Figure 32** demonstrates, in pooling effects from eight different studies we find that certification lowers wages of workers in agricultural production (SMD -0.26, 95%

[77] As the (purposeful) vagueness of the language employed here indicates, the issue is further complicated by disagreements over basic definitions. There is no single accepted definition of what business scale constitutes a smallholder farmer as opposed to a plantation and studies differ in their respective definitions.
[78] And Cramer et al (2014) control for business scale.

CI from -0.46 to -0.06). This effect is statistically significant (p=0.012).[79] Heterogeneity is substantial (I^2=86.3%). The meta-analysis concerns only wages, and we have very limited information on working conditions. In addition to their wage effect estimate, Ehlert et al (2014) also report that Kenyan workers in certified fruit and vegetable production receive more training than workers in non-certified production. Schuster and Maertens (2014) find that workers in certified Peruvian fruit and vegetable production companies tend to be employed for longer, indicating greater job security. Among studies that we could not extract effect size estimates from, van Rijn et al (2016) report no statistically significant effects on wages for workers on banana plantation in both Colombia and the Dominican Republic.

Figure 32: Forest plot for wages



Diagnostics: Heterogeneity chi-squared = 50.94 (d.f. = 7) p = 0.000; I-squared (variation in ES attributable to heterogeneity) = 86.3%; Estimate of between-study variance Tau-squared = 0.0637; Test of ES=0: z= 2.51, p = 0.012.

It is notable though that all statistically significant effects come from Cramer et al (2014). Repeating the analysis without the four studies provided by Cramer et al (2014) yields a non-significant pooled effect (SMD 0.012, with a 95%-confidence interval ranging from 0.00 to 0.02). This result is in turn largely driven by Dragusanu (2014), which gains in weight following the exclusion of Cramer et al (2014), which had relatively large sample sizes. Heterogeneity falls to zero.

[79] As laid out in Section 3.3.6. a single report may contain data from several different studies, i.e. unique datasets. In this instance Cramer et al (2014) contains data from four separate studies and therefore appears four times in the forest plot.

As for previous outcomes, **Figure 33** breaks the effects up by certification scheme. Fairtrade is the only scheme to produce on overall statistically significant result, driven by negative results reported by Cramer et al (2014). Both Fairtrade and GlobalGAP show a wide dispersion of effect size estimates, though the heterogeneity is larger for Fairtrade (see **Table 12** below).

Figure 33: Forest plot for wages by certification scheme

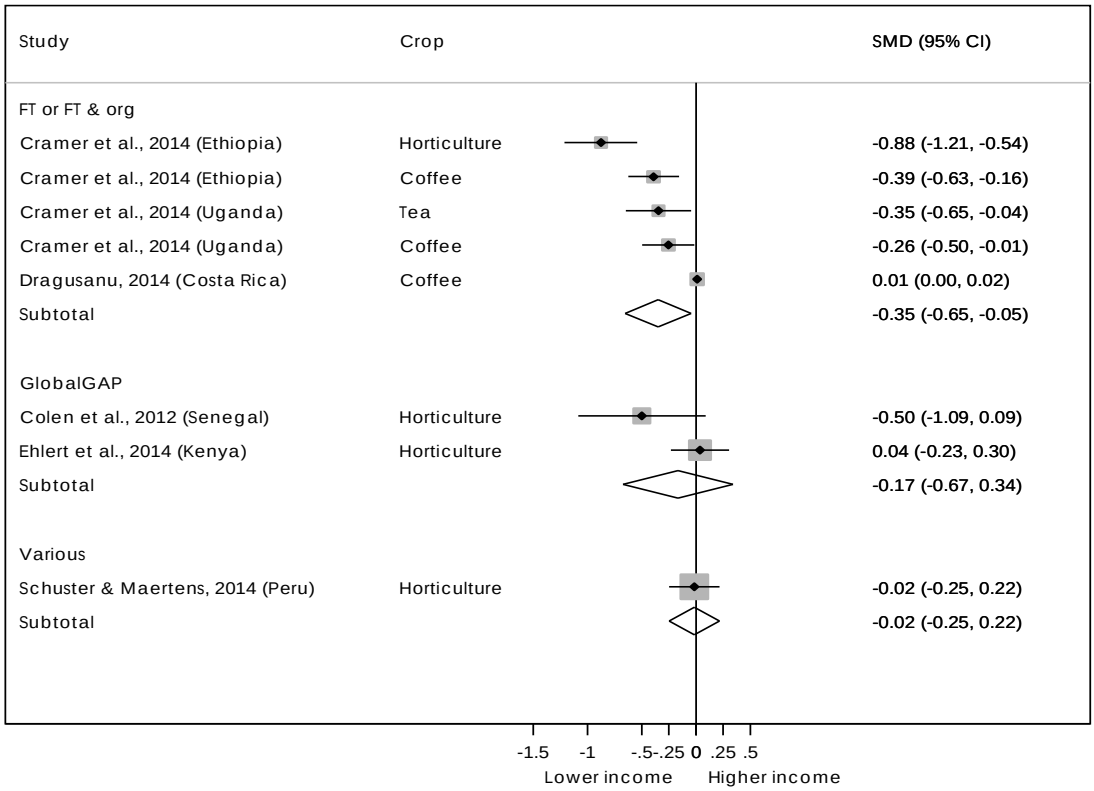


Table 12: Heterogeneity statistics for moderator plot

Category	Q	T^2	I^2	p-value for Q	# of studies
FT or FT & org	47.99	0.1041	91.7%	0.00	5
GlobalGAP	2.63	0.0886	62.0%	0.105	2
Various	0	0	0	-	1

Conducting the analysis separately for studies that show high (Dragusanu, 2014; Ehlert et al, 2014), moderate (Cramer et al, 2014, Colen et al, 2012) and low risk of bias (Schuster and Maertens, 2014) gives statistically non-significant results for the high risk and low risk studies, while the moderate risk of bias group continues to produce a statistically significant negative result. This is again driven by Cramer et al (2014). Looking only at studies published in peer-reviewed journals, i.e. Colen et al (2012), Dragusanu (2014), and Ehlert et al (2014), gives mildly negative and not statistically significant pooled effect. All studies included for this outcome were fully independently financed and there is thus no difference between them in this regard.

Table 13: Sensitivity analysis for wages

Category	SMD	CI lower bound	CI upper bound	Q	T^2	I^2	p-value for Q	# of studies
All studies	-0.26	-0.46	-0.06	50.94	0.0637	86.3%	0.00	8
Risk of bias								
High	0.012	0.000	0.024	0.03	0.00	0.0%	0.862	2
Moderate	-0.450	-0.661	-0.240	9.24	0.0312	56.7%	0.055	5
Low	-0.015	-0.248	0.217	0	0	0	-	1
Peer-reviewed publication								
Peer-reviewed journal	-0.011	-0.156	0.133	2.94	0.0073	31.9%	0.230	3
Other publication	-0.360	-0.609	-0.111	17.75	0.0618	77.5%	0.001	5

4.4.6 Total household income

Up to now we have dealt with intermediate outcomes. Following the causal chain we now look at endpoint outcomes. Returning to farmers, we noted in Section 4.4.4 that a central concern for most certified agricultural producers was the gain in income they could expect from engaging in such production.[80] However, in that section we looked at the income gained directly from the production and sale of certified commodities. While an important indicator, this is not the same as total household income. Consumption at household level is largely determined by total household income (abstracting from borrowing, savings, gifts and subsistence production) and this can have a complicated relationship with income from certified production. For instance a rise in income from certified production may unambiguously lead to a rise in total household income. But certification frequently involves both financial and opportunity costs. Certification may demand additional labour inputs, or simply time to attend cooperative meetings and trainings, which may not be available for the production of other crops or for other income generating measures. Certification may thus also affect the amount of wage work undertaken by household members, both through increased specialisation and through additional labour needs on the family farms. Moreover, the effect of income from certified production also depends on the share of household income that is derived from the certified commodity. The relation between total household income and income from certified production is further complicated by possible shifts in the intra-household distribution of labour in

[80] This is of course not to denigrate other intrinsically worthwhile goals such as environmental protection or healthier production practices.

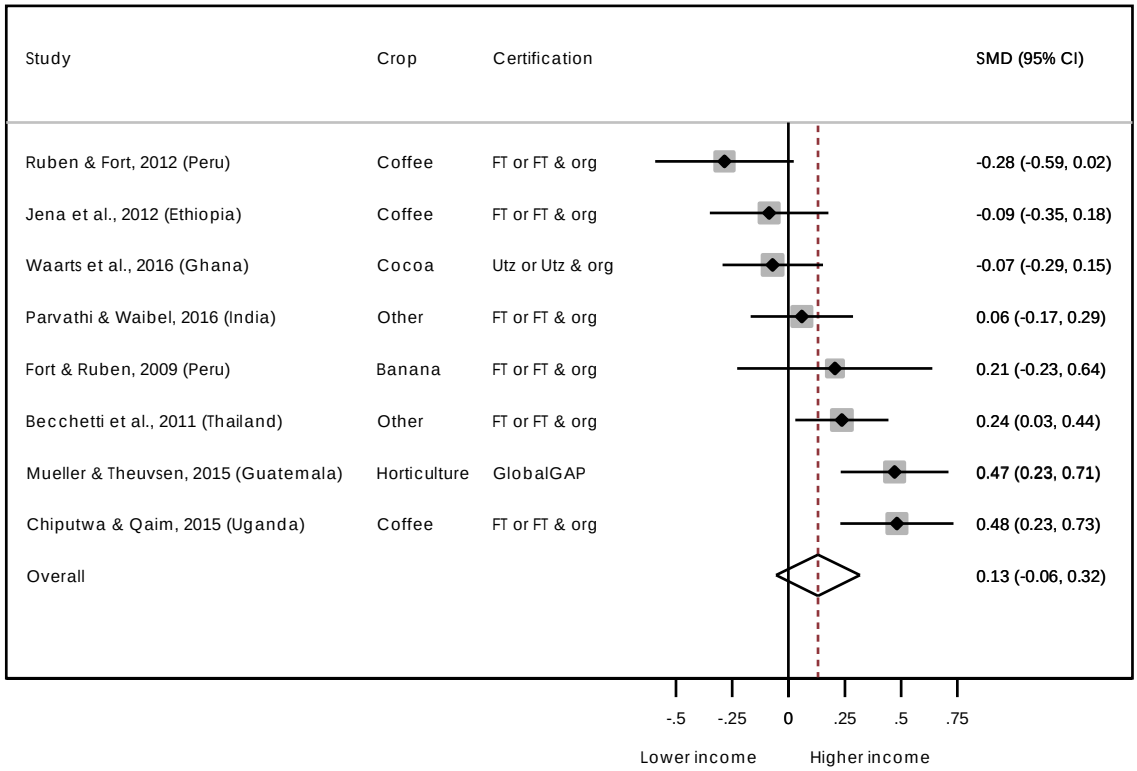
response to certification.[81] It is therefore not clear *a priori* that an increase in income from certified production does in fact lead to higher income at household level.

In this section we therefore look at the effect of certification at total household income. Studies included here used surveys to determine the change in income at the level of the household. Unlike the other studies considered, Chiputwa and Qaim (2015) do not measure the effect on income but rather consumption. We have nonetheless included their study in the analysis, as income and consumption are closely related at household level, so much so, that household surveys commonly gather data on consumption as a proxy for income. Household income is notoriously difficult to measure with any real accuracy and none of the included studies provided enough information on the way they arrived at their household income measures. The effect size estimates used here may thus be subject to substantial measurement error in the underlying reported effect estimates.

With these caveats in mind we find that the pooled effect from eight studies is an increase in total household income as a result from certification (SMD 0.13, 05% CI from -0.06 to 0.32) (see **Figure 34)**. The pooled effect is however not statistically significant (p=0.17). Once again, there is substantial heterogeneity across studies (I^2=76.2%). Jena et al (2012) also provided an estimate of income per capita, but we have included their estimate of total household income instead, as this is more closely comparable to the effects measured by the other studies. Interestingly, this heterogeneity is present despite the fact that six out of the eight studies examined Fairtrade certification. Strikingly, the most negative – albeit not statistically significant – effect (SMD -0.28, 95% CI from -0.59 to 0.02), provided by Ruben and Fort (2012) and the most positive effect (SMD 0.48, 95% CI from 0.23 to 0.73) from Chiputwa and Qaim (2015), both look at Fairtrade certified coffee producers, albeit on different continents. The confidence intervals of these effect estimates do not overlap, making it very unlikely that this is a purely statistical phenomenon. This illustrates the point we make repeatedly across this review, namely the importance of contextual factors in the impact of such schemes.

[81] Which may also affect the intra-household distribution of income.

Figure 34: Forest plot for total household income



Diagnostics: Heterogeneity chi-squared = 29.39 (d.f. = 7) p = 0.000; I-squared (variation in ES attributable to heterogeneity) = 76.2%; Estimate of between-study variance Tau-squared = 0.0537; Test of ES=0: z= 1.37, p = 0.170.

Excluding Chiputwa and Qaim (2015) does not change the results. Rerunning the analysis only with studies that report income rather than consumption still produces a positive pooled effect (SMD 0.079, with a 95%-confidence interval from -0.11 to 0.26), meaning that the effect is also not statistically significant (p=0.401). Heterogeneity remains broadly similar (I^2=71.9%).

Separating the results out by certification schemes, as done in **Figure 35**, shows that there is no evidence that any scheme has a statistically significant impact, apart from GlobalGAP, which however is based on just one study. As for other outcomes discussed the wide variety of effects within a single scheme is striking. Studies analysing Fairtrade for instance encompass two negative effect size estimates along with two positive ones, though only the latter two are statistically significant. As already noted, among the Fairtrade effects, the two negative estimates and the largest positive effect size all deal with coffee. Heterogeneity statistics for the plot are shown in **Table 14** below.

Figure 35: Forest plot for total household income by certification

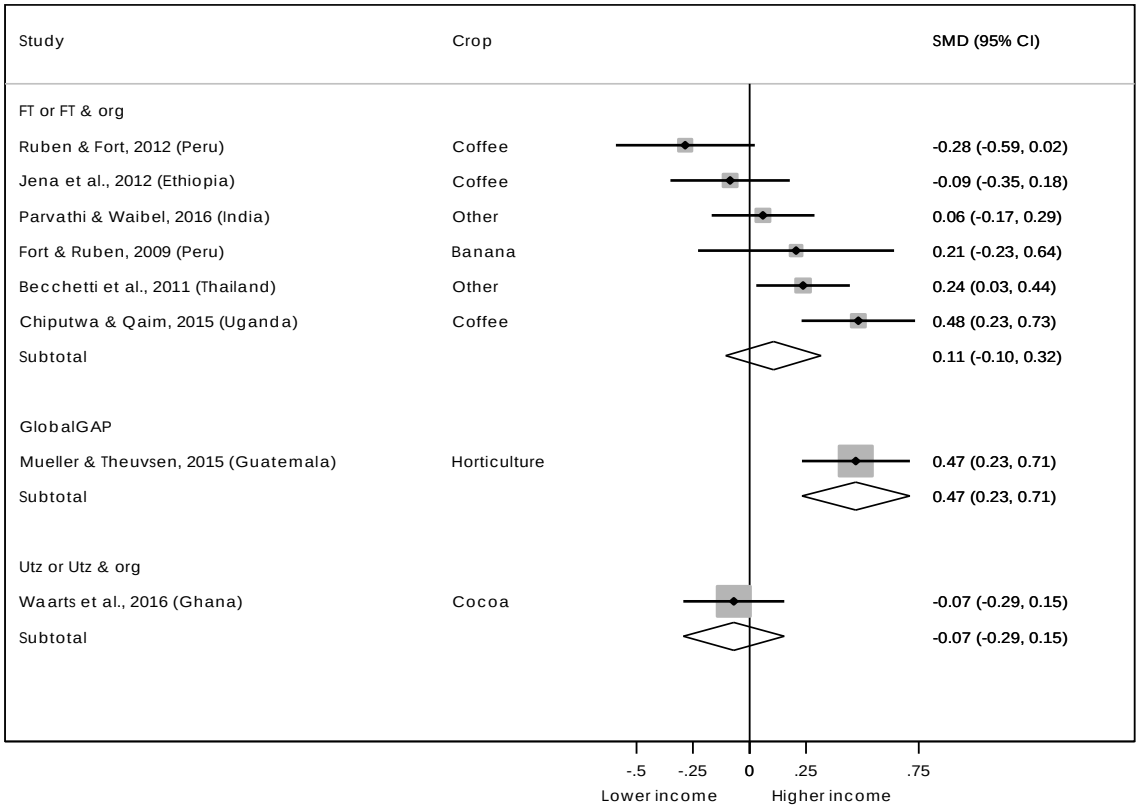


Table 14: Heterogeneity statistics for moderator plot

Category	Q	T^2	I^2	p-value for Q	# of studies
FT or FT & org	18.58	0.0494	73.1%	0.002	6
GlobalGAP	0	0	0	-	1
Utz or Utz & org	0	0	0	-	1

The null result for the pooled effect is very robust, as shown in **Table 15**. Results remain positive but not statistically significant when comparing high risk of bias (Jena et al, 2012; Waarts et al, 2016; Parvathi and Waibel, 2016; Becchetti et al, 2011) and moderate risk of bias studies (Ruben and Fort, 2012; Fort and Ruben, 2009; Mueller and Theuvsen, 2015; Chiputwa and Qaim, 2015). Fully independently financed studies (Jena et al, 2012; Parvathi and Waibel, 2016; Mueller and Theuvsen, 2015; Chiputwa and Qaim, 2015) and not fully independent studies (Ruben and Fort, 2012; Fort and Ruben, 2009; Waarts et al, 2016; Becchetti et al, 2011) also produce positive and not statistically significant pooled effects across both groups. The only notable difference emerges between studies published in peer-reviewed journals and those published elsewhere. Studies from peer-reviewed journals (Waarts et al, 2016; Fort and Ruben, 2009; Mueller and Theuvsen, 2015; Chiputwa and Qaim, 2015) produce a statistically non-significant pooled effect centred on zero, while the other studies return a positive effect that is only just not statistically significant.

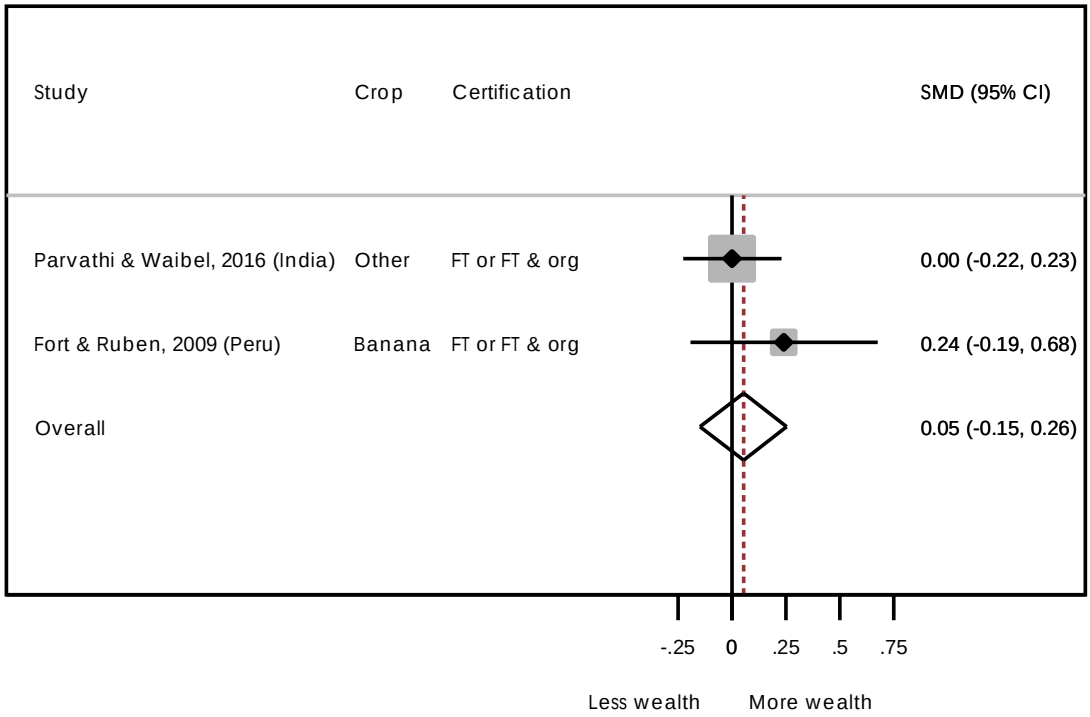
Table 15: Sensitivity analysis for total household income

Category	SMD	CI lower bound	CI upper bound	Q	T^2	I^2	p-value for Q	# of studies
All studies	0.13	-0.06	0.32	29.39	0.0537	76.2%	0.00	8
Risk of bias								
High	0.045	-0.107	0.197	5.27	0.0103	43.0%	0.153	4
Moderate	0.228	-0.133	0.590	17.97	0.1112	83.3%	0.000	4
Study independence								
Fully independent	0.233	-0.045	0.511	15.42	0.0646	80.5%	0.001	4
Not fully independent	0.019	-0.223	0.261	9.25	0.0398	67.6%	0.026	4
Peer-reviewed publication								
Peer-reviewed journal	0.002	-0.209	0.212	8.66	0.0299	65.3%	0.034	4
Other publication	0.274	-0.023	0.572	14.39	0.0709	79.2%	0.002	4

4.4.7 Assets/wealth

As noted in the overview, with the partial exception of schooling, we have much less data on endpoint outcomes and the results in this section and the next are thus purely illustrative. Our first endpoint outcome is the effect of certification on the wealth status of producers. As with household income there are substantial practical difficulties involved in compiling and accurately pricing asset indices and other measures of wealth (see Johnston and Abreu, 2016 for a recent discussion). Neither of the two studies considered here provided much information on how this information was collected, nor on how their indices were put together, and there may be substantial measurement error. In part the lack of data may be related to such measurement issues, as researchers may be reluctant to try and provide estimates based on noisy data.

Figure 36: Forest plot for wealth



Diagnostics: Heterogeneity chi-squared = 0.92 (d.f. = 1) p = 0.337; I-squared (variation in ES attributable to heterogeneity) = 0.0%; Estimate of between-study variance Tau-squared = 0.0000; Test of ES=0: z= 0.53, p = 0.598.

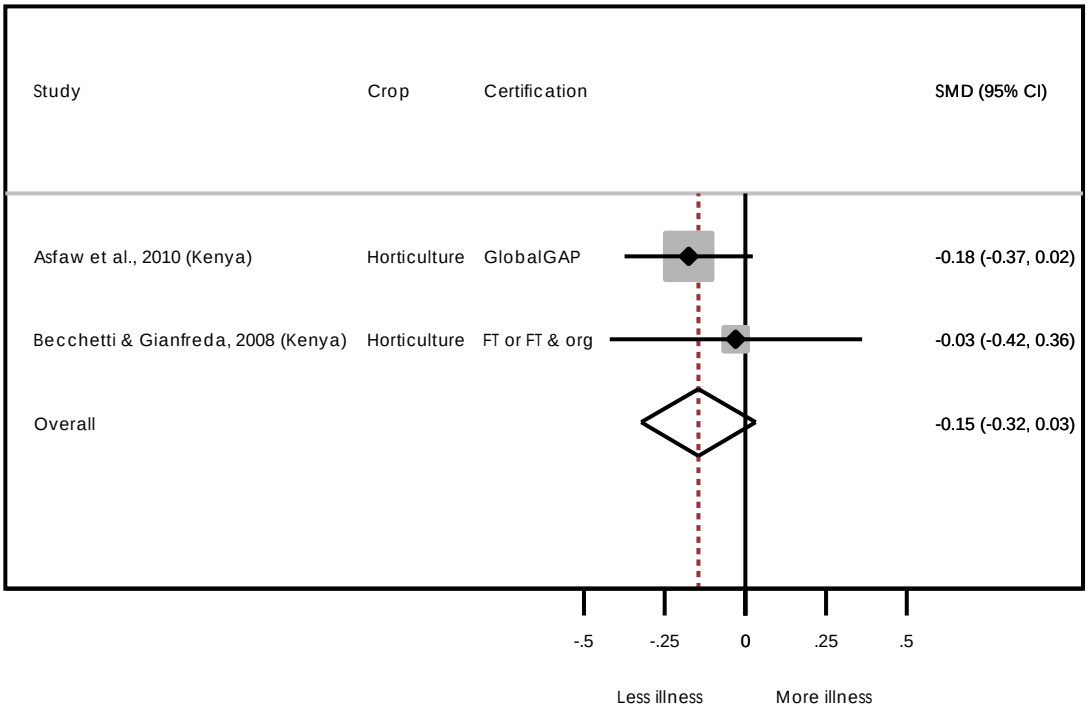
Due to data limitation we pool effect sizes from just two studies in **Figure 36** above. With so few studies the pooled effect has only limited value. The joint effect is a very slight increase in wealth among certified producers (SMD 0.05, 95% CI from -0.15 to 0.26), but the effect is not statistically significant (p=0.598). As there are just two studies and their confidence intervals overlap the standard measures of heterogeneity are zero, while the Chi-squared statistic is too instable to be given much credence. Both studies provide statistically insignificant results, leading us to conclude that the very limited evidence we have shows no increase in wealth. However, Parvathi and Waibel (2016) has a high risk of bias, while Fort and Ruben (2009) who provide a larger, though still not statistically significant effect size estimate, have only moderate risk of bias. Of course, future reviews will hopefully be able to rerun this analysis with a wider range of studies reporting on a more certification schemes. Due to the limited number of studies available, we did not conduct moderator or sensitivtiy analysis for this outcome.

4.4.8 Illness

Health, defined here simply and reductively as the absence of illness, is a central determinant of individual welfare and of human development more broadly (see for instance Deaton, 2013). As for assets above, we have just two studies that give evidence on health, and the conclusions of this section are therefore only illustrative. The two studies we have that look at illness both report that certified producers use less pesticides than their non-certified counterpart. Both directly relate reductions in

illness to lower pesticide use, although Becchetti et al (2008) also emphasise that another causal chain runs through higher incomes. Asfaw et al look at GlobalGAP which places limits on the types and amounts of pesticides that can be applied to plants, while the subjects in Becchetti et al's study are double-certified with both Fairtrade and organic certifications. It stands to reason that the reduction in pesticide use in the latter case might be more to do with the organic certification.

Figure 37: Forest plot for illness



Diagnostics: Heterogeneity chi-squared = 0.42 (d.f. = 1) p = 0.515; I-squared (variation in ES attributable to heterogeneity) = 0.0%; Estimate of between-study variance Tau-squared = 0.0000; Test of ES=0: z= 1.61, p = 0.106.

The pooled effect that emerges from the meta-analysis indicates a slight reduction in illness, that is, an SMD of -0.15 with a 95%-confidence interval ranging from -0.32 to 0.03, meaning that the effect is not statistically significant (p=0.106). Standard measures of heterogeneity are zero. As both studies measure incidences of illness, a negative effect indicates a reduction in illness, which of course is a good thing. The very limited evidence we have here however leads us to conclude that certification has no statistically significant effects on illness. Both studies are rated as having a high risk of bias, which does not serve to increase confidence in these findings. As above, the limited number of studies makes moderator and sensitivity analysis unnecessary for this outcome.

4.4.9 Schooling

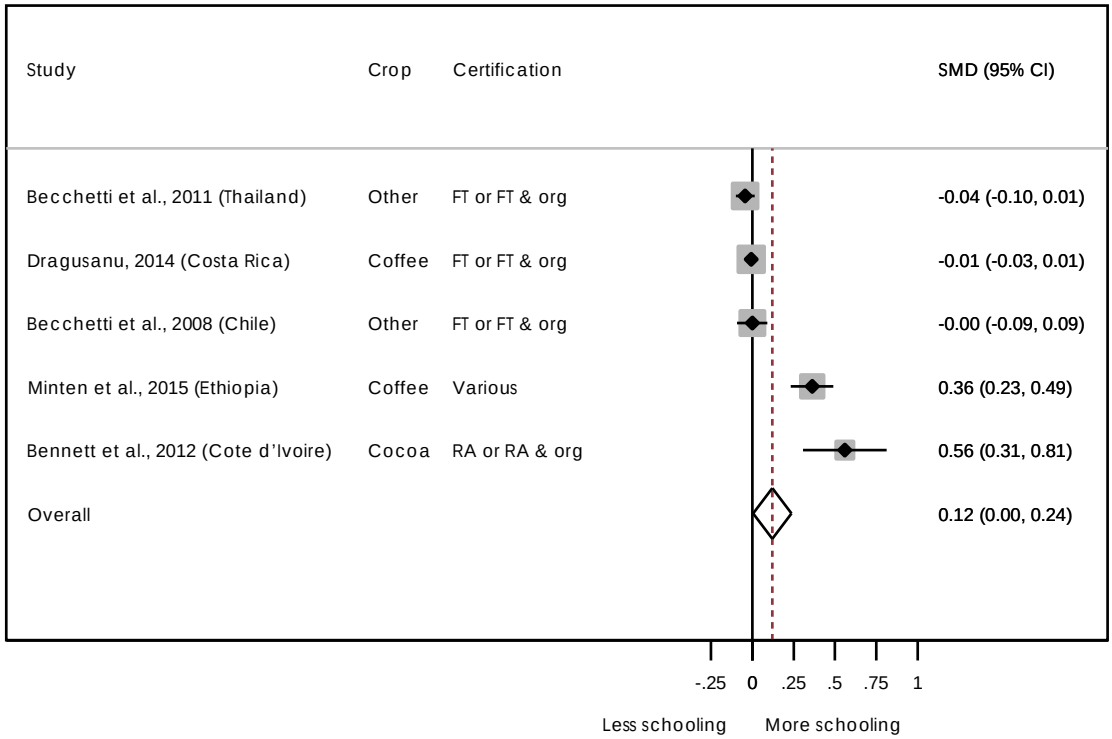
Of the endpoint outcomes we consider, the best evidence we have comes on the effect of certification on schooling. All studies compiled here look at the effect of certification on the school attendance rates of children who live in households that

produce certified commodities. Again, the relationship is not necessarily straightforward. An obvious point is child labour. For instance, an increase in income from certified production could lead to higher household income, and the increase in income could make the direct economic contribution of children to the household unnecessary and/or provide the funds necessary to send them to school. For instance, Minten et al (2015) find that certification reduces the use of child labour. On the other hand however, such an increase in income raises the value of labour inputs into the certified commodity and could lead to more child labour, either directly in production or substituting for adult labour in the household thus freeing adults to produce valuable commodities. Beyond putting limits on the use of child labour, many CS do not contain direct mechanisms by which to increase the school attendance rates of children. A partial exception is Fairtrade, as the social premium paid to certified cooperatives can be used to build schools, for instance. Even here there is no requirement for funds to be spent on education though. This depends on decisions made by the governing bodies of POs or plantations.

We pooled effect sizes on schooling from five studies. From these we estimate an increase in school attendance as a result of certification (SMD 0.12, 95% CI from 0.00 to 0.24). The pooled effect is statistically significant (p=0.041). The results show very high levels of heterogeneity (I^2=92.3%). Looking at **Figure 38**, this is not surprising. While Minten et al (2015) and Bennett et al (2012) show statistically significant positive effects, the three other studies have null results centred closely around zero.

For schooling we do not require a separate forest plot to see how effects differ across schemes. All of the estimates of Fairtrade impact are closely centred on zero, and hence are statistically not significant, while the single estimate for RA is strongly positive and statistically significant. Minten et al (2015) combine certification including Fairtrade, Utz and RA, in such a way as to make the attributon of effect to a single scheme impossible, but they report a statistically signifcant positive effect for being certified.

Figure 38: Forest plot for schooling



Diagnostics: Heterogeneity chi-squared = 52.12 (d.f. = 4) p = 0.000; I-squared (variation in ES attributable to heterogeneity) = 92.3%; Estimate of between-study variance Tau-squared = 0.0142. Test of ES=0: z= 2.04, p = 0.041.

Sensitivity analysis, presented in **Table 16**, shows clear differences between groups when studies are stratified by their risk of bias rating. The three statistically non-significant results were produced by studies with high risk of bias ratings (Becchetti et al, 2008; Becchetti et al, 2011; Dragusanu, 2014), while the two positive and statistically significant results came from studies with moderate risk of bias (Minten et al, 2015; Bennett et al, 2012). This strengthens our belief in an underlying positive effect of certification on schooling. When viewed as separate groups, both fully independently financed studies (Dragusanu, 2014; Minten et al, 2015) and not fully independently financed studies produce pooled effects that are positive but not statistically significantly different from zero. Looking separately at studies published in peer-reviewed journals (Becchetti et al, 2011; Dragusanu, 2014) we find that these studies yield a pooled estimated effect of almost zero, while studies published elsewhere give a positive effect. Neither are statistically significant. Given the small number of studies, the results of this sensitivity analysis should be interpreted with caution.

Table 16: Sensitivity analysis for schooling

Category	SMD	CI lower bound	CI upper bound	Q	T^2	I^2	p-value for Q	# of studies
All studies	0.12	0.00	0.24	52.12	0.0142	92.3%	0.00	5
Risk of bias								
High	-0.010	-0.030	0.009	1.48	0	0.0%	0.476	3
Moderate	0.429	0.245	0.614	1.89	0.0093	47.1%	0.169	2
Study independence								
Fully independent	0.172	-0.188	0.532	30.89	0.0654	96.8%	0.000	2
Not fully independent	0.118	-0.078	0.315	20.95	0.0251	90.5%	0.000	3
Peer-reviewed publication								
Peer-reviewed journal	-0.015	-0.046	0.016	1.44	0.0002	30.3%	0.231	2
Other publication	0.292	-0.028	0.612	30.87	0.0729	93.5%	0.000	3

4.4.10 Summary

In summary, we find that the available evidence does not give a clear picture of the impact – or lack thereof – of certification schemes. For yields we synthesised five studies and found a reduction in yields (SMD -0.42, CI from -1.23 to 0.39) for certified farmers, although the effect is not statistically significant (p=0.312). For price, four studies provide our pooled estimate of an increase in the price received (SMD 0.28, 95% CI from 0.06 to 0.47) and the effect is statistically significant (p=0.005). We have the most evidence for income from certified production. Ten studies return a pooled effect size showing a rise in income for certified producers (SMD 0.22, 95% CI from 0.03 to 0.41). The overall effect is statistically significant (p=0.021). On wages however, across eight studies we find that certification lowers wages of workers in agricultural production (SMD -0.26, 95% CI from -0.46 to -0.06). This effect is statistically significant (p=0.012). However, this result is largely driven by four studies conducted by the same research team. Removing these studies produces a null result that is not statistically significant. Possibly one of the most important outcomes for farmers is the change in total household income as a result of certification. Here eight studies show a combined increase in total household income as a result from certification (SMD 0.13, 95% CI from -0.06 to 0.32). The pooled effect is however not statistically significant (p=0.17). The evidence base is weakest for effects on wealth and illness, as we have evidence from just two studies for each. For wealth the joint effect is an increase in wealth among certified producers (SMD 0.05, 95% CI from -0.15 to 0.26), but the effect is not statistically significant (p=0.598). The meta-analysis for illness finds a pooled effect showing a decrease in illness amongst

certified producers (SMD -0.15, 95% CI from -0.32 to 0.03), though again the effect is not statistically significant (p=0.106).[82] Finally, for schooling we estimate an increase in school attendance as a result of certification (SMD 0.12, 95% CI from 0.00 to 0.24). The pooled effect, which comes from five studies, is statistically significant (p=0.041).

While the evidence in hand points largely towards findings that are not statistically significant, the evidence base is also too thin in most cases to have great confidence in these findings. This is in large parts the result of the absence of impact evaluation that met the inclusion criteria for this review. However, we have substantially more data from qualitative research, looking at barriers and facilitators to impact, to which we turn next. In Section 4.6 we then combine the findings from the quantitative and qualitative syntheses.

4.5 Qualitative synthesis of results

This section presents the results of the synthesis of findings from the qualitative studies included for RQ2. A table summarising the main findings (**Table 17**) is provided at the end of this section for readers to see the most salient barriers, enablers and contextual factors. The synthesis is based on 138 studies across 114 individual reports that meet the inclusion criteria for RQ2, as outlined in Section 3.1. In the interest of maintaining focus, the list of included studies, along with a descriptive summary for each study, is made available in Annex B rather than the main text. Each included study contributed differently to the qualitative synthesis and to the evidence themes we develop below. The contribution of each study to the main themes is documented in Annex G.

Drawing on Thomas and Harden (2008), we used a thematic synthesis approach in three stages (detailed coding, generation of descriptive themes, generation of analytic themes) to synthesise data from the included studies. A detailed account of the data synthesis methods can be found in Section 3.3, while the coding tools used to synthesise data from included studies are provided in Annex D. Each thematic sub-section of this qualitative synthesis also indicates the number of studies that contributed substantive evidence to each specific theme, which can give a sense of how much evidence we found for different themes as well as where the main focus lies within the literature.

The synthesis is presented using the hypothesised synthetic theory of change (ToC) as an overall framework, into which emerging issues are incorporated. After a preliminary review of the evidence of studies eligible for inclusion with regard to RQ2, we decided to organise the evidence on barriers and facilitators and contextual factors as follows: First we present a synthesis of findings related to the implementation dynamics of CS, which focuses on certification inputs, certification-related costs, monitoring and auditing, spill-over and unintended effects, as well as the dynamics of multi-certification. This is followed by a synthesis of findings related to distributional dynamics, in particular entrance into, and adoption of, certification

[82] As we are dealing with illness, a negative SMD indicates a desirable result in this case.

standards along the lines of wealth and gender. Finally, we synthesise the findings related to contextual barriers and facilitators, focusing on producers and Producers Organisations (POs), plantations, markets, institutional environment, and socio-economic context.

As it can be observed from the risk of bias assessment (Section 4.2), the studies included present important methodological limitations. Biases in terms of study focus, such as over- or under-represented geographical areas and commodities should also be considered in the interpretation of the results. Finally, we remind the reader that the qualitative synthesis does not seek to establish whether CS work or not; this was the objective of the quantitative synthesis. Instead, it seeks to complement and explain the findings of the quantitative synthesis by illuminating how, when, why and for whom CS may or may not work. For this reason, as discussed in Section 3.3.2, descriptive data from included studies that only reported on effects without providing any insights on implementation or distributional dynamics, or how the context can shape these effects were not included in this synthesis. It is important to note, therefore, that as primary studies tend to focus more on barriers and less on facilitators, this is inevitably reflected in the final synthesis. For all the above reasons, the reader is invited to interpret the results with caution, as an overall account of the existing literature on CS, but not as definite account of the reality of CS, which, as this review has shown, is complex, dynamic and highly context-sensitive.

4.5.1 Implementation dynamics

This section presents a synthesis of findings on implementation dynamics of CS, drawing on 127 studies. For each sub-section the number of contributing studies is reported in brackets.

Targeting and (self-) selection of participants

Although adequate selection of farmers according to aims (of the CS) is a key assumption of the ToC, the review identifies a striking lack of systematic reporting on targeting processes of CS, such as the decision making mechanisms of implementing actors (i.e. certification bodies and NGOs) regarding geographical location and targeted population. Nevertheless, a number of studies provide insights regarding the way producers, POs and plantations enter certified markets and adopt certification standards. Findings are presented along the lines of (differences in) wealth and gender, the two pillars we identified as highly significant.

Wealth and resources (n=35)

The synthesis indicates that there may be important barriers to entry in a certification process. The adoption of standards required by a CS often depends on the capacity of POs, producers and plantations to bear the extra costs related to certified production, i.e. the implementation of standards, costs of certification/registration or follow-up costs (see Section 4.5.1.3). This capacity is highly dependent on the wealth and resources available to producers (Asfaw et al, 2009b), but also on their capacity to obtain external support to finance the certification process by aid providers, buyers, or partnerships of actors (for instance Dowdall, 2012; Rotter, 1999; Pongratz-Chander, 2007, among others). The following section presents a synthesis of

findings from included studies related to the importance of producers' wealth and resources in becoming certified. Evidence related to the importance of external support to finance the certification process is presented in below in the sections on ‘Support to strengthen POs’ and ‘POs context: management, relationship with producer and with buyers’.

The synthesis of findings on the role of wealth and resources in the certification process points to the following. First, it is suggested that increased costs of certified production, mainly driven by higher labour costs of organic production and stricter quality criteria, discourage producers with less resources to join certified POs (Milford, 2014; Jaffee, 2006; Abarca-Orozco, 2015), but also directly exclude the producers who cannot afford compliance with higher quality standards (Getz and Shreck, 2006; Shreck, 2002; Smith, 2010). Another decisive factor appears to be producers’ financial ability to withstand important payment delays in certified POs (Donovan and Poole, 2014; Donovan and Poole, 2014b; Abarca-Orozco 2015; Dowdall, 2012; Rotter, 1999; Valkila, 2009; Valkila and Nygren, 2008; Ronchi, 2002; Bagama et al, 2014). Such ability is reported to be related to diversification of sources of income rather than exclusive dependence on the certified crop. As a result other sources of advantage, such as higher education or capital in form of productive assets play an important role (Dowdall, 2012). A lack of liquidity can also lead to producers leasing their lands and cashable assets to local traders, therefore compromising future harvests and undermining their participation in certified markets (Makita, 2011).

Production capacity, land size, and the degree of market integration are also reported to influence participation in CS (Subervie and Vagneron, 2013; Cofre et al, 2012; Kariuki, 2014). For example, Cofre et al (2012), reporting on GlobalGAP, provide evidence that certified producers tend to be larger, and more prosperous, with higher educational levels, as well as greater involvement in producer associations. On the other hand, the combination of social-organic standards is reported to be less attractive for producers with smaller plots and fewer resources (Abarca-Orozco, 2015; Milford, 2014; Chiputwa, 2015).

Selection based on size may also occur at the PO or factory level. For example, some certified POs are reported to formally exclude very small producers (Nelson and Martin, 2013; Laroche et al, 2012; Staib, 2012), the rationale being that they require more extension services while producing relatively small quantities, therefore not generating sufficient turnover for the PO to support them (Laroche et al, 2012; Staib, 2012). Baker (2014), reporting on Rainforest Alliance tea factories in Argentina, states that factories were actually selecting larger farms that were already close to meeting all the requirements in order to minimise the costs of their incorporation in the production chain. Finally, Fairtrade certification is reported to clash with the more diversified agriculture of smaller farmers who cannot afford to divert more efforts and resources on the certified crop (Makita, 2011). Household size appears to matter as well, as reported evidence indicates that Fairtrade-organic producers have more household members able to work, a fact possibly linked to the increased labour requirements of certified production (Abarca-Orozco, 2015; Milford, 2014; Chiputwa, 2015). Education and literacy skills also appear to facilitate

participation in certified POs due to the paperwork requirements at both farm and PO levels, while language barriers and illiteracy can hinder participation (Bacon, 2005; Lyon et al, 2010; Laroche et al, 2012; Pollack, 2006).

Moreover, evidence suggests that early adopters are larger and better established farmers in terms of land tenure, farming experience, and length of local residence (Laroche et al, 2012; Moberg, 2008). Newcomers, on the other hand, who tend to be poorer and more marginalised farmers, are reported to face difficulties in joining certified POs, ending up instead on waiting lists (Moberg, 2005), as POs, particularly successful ones, may become saturated and not able to accommodate more members (Mendez, 2002). This trend can be accentuated by certification policy on audits, as in the case of Rainforest Alliance group certification Bakker (2014).[83]

This can also apply at the PO level due to limited demand for certified products. For example, veteran Fairtrade coffee POs are reported to dominate the market while new ones face difficulties in establishing new, long-standing relationships with Fairtrade buyers (Valkila and Nygren, 2008). Additionally, Pongratz-Chander (2007) suggests that Fairtrade organisations, but also development agencies and lending institutions, tend to work with POs that are stable, well-established, have proven to be democratic, and can be reliable providers in terms of volumes and quality. This means that it can be challenging for young POs with less experience not only to enter certified markets, but also to receive the guidance and financial support needed to adopt standards and access such markets (i.e. Kariuki, 2014; Beall, 2012).

Finally, wealth and size also matter at the company level. Evidence from flower plantations in Ecuador, for instance, suggests that only the wealthiest companies were able to afford experimenting with certified production and to have the capital and infrastructure to comply with quality standards (Raynolds, 2012; Schelly, 2011), while Staricco and Ponte (2015) report that it is mostly the already dominant Argentinian wineries that benefit from Fairtrade.

Overall, the synthesis of findings suggests that there are important and systematic pre-existing differences in wealth and resources between certified producers, POs and plantations and non-certified organisations or newcomers. Such differences are crucial not only in terms of impact attribution and correction of (self-) selection bias, but also in terms of certification reach, as despite CS claims about improving trading conditions for the "small-scale" and "economically disadvantage producers" (i.e. WFTO, 2017; Fairtrade International, 2017) and addressing poverty of "smallholder" and indigenous farmers (i.e. Utz, 2014; Rainforest-Alliance, 2014),[84] it appears that

[83] Bakker (2014) reports that the policy of conducting a full audit (all 99 criteria must be checked again, as in the first audit) when more than 10% of the total land area of the previous year is added to the group certification, resulted in Rainforest Alliance tea factories/groups limiting the total land size of the new participants to add less than 10% of its total area each year and avoid a full audit, hence narrowing the chances of non-certified producers to join.
[84] https://www.utz.org/better-business-hub/strengthening-your-reputation/prosperity-for-cocoa-farmers-just-around-the-corner/
http://www.rainforest-alliance.org/articles/rainforest-alliance-certified-cocoa

CS are not generally able to reach and deliver benefits to the farmers that need them the most.

Gender (n=22)

Gender equality is an important aspiration in some CS. Therefore a key question is whether CS contribute to women's empowerment. However, this primarily depends on the extent to which women do have access to CS interventions and their benefits in the first place. The majority of available studies focused on women's participation in certified POs report limited female participation, suggesting that certification-related gender equity programmes or strategies have not been able to offset the socio-cultural and organisational barriers that women commonly face (Bacon, 2005; Bergeron, 2012; Ellery, 2010; Hanson et al, 2012; Mendez, 2002; Pollack, 2006; Pongratz-Chander, 2007; Smith, 2010; Waarts et al, 2016; CESU, 2012; TWIN, 2013). Women's weak participation in certified POs is explained by the following reasons.

First, female producers may find it difficult to divert attention and time from their households or their farms to attend certification-related activities (Hanson et al, 2012; Nelson and Martin, 2013; TWIN, 2013; Stenn, 2015). In fact, Hanson et al (2012:171), reporting on Fairtrade coffee producers from Nicaragua, highlight that while women's workloads from domestic and care-giving work remain unaffected, certification may result in an 'onerous double burden of work'. Second, lower education, lack of skills and knowledge are also keeping women from participating in certified POs, and particularly from accessing leadership positions (Sutton, 2014; Terstappen, 2010; Pongratz-Chander, 2007).

Third, the socio-cultural context appears to play a major role. Unequal gender relations and violence (i.e. machismo) are commonly reported as a barrier to female participation in certified initiatives, as women often face mobility restrictions, disapproval regarding their choices or disbelief regarding their abilities, while they tend to be excluded from the economic and social benefits of certified production, even though they significantly contribute with their work (Pollack, 2006; Sen, 2009; Pongratz-Chander, 2007; Fairtrade, 2015). For example, travelling restrictions for women, (i.e. out of fear of harassment, or lack of childcare), are reported to effectively limit their capacity to participate in meetings and training (Lyon et al, 2010; Pollack, 2006; Stenn, 2015; Stathers and Gathuthi, 2013), while lack of financial support to travel is also reported (Stenn, 2015). Additionally, socio-cultural norms and perceptions can prevent women's participation, particularly after marriage (TWIN 2013; Sen, 2009; Sutton 2014).Overall, it seems that it is particularly poorer women with heavier workloads who are more likely to be excluded from participation, although women from wealthier households are also reported to be restricted (Sen, 2009).

Limited cases of increased female participation are also reported but should be carefully interpreted. Records of female membership, for instance, can be misleading as women may be members only on paper just to enable better household access to credit from the PO, or to increase voting rights, as Ronchi (2002) reports. There are

some unexpected factors, however, that do appear to enhance female participation in certified POs. Male migration is one, as male 'operators' who migrate pass the 'operator' status to their wives,[85] who are able to divert more time and attention to certified production, due to reduced housework resulting from the absence of the husband (and sometimes also the children) (Lyon et al, 2010). The formal recognition in CS also makes visible the role of women in certified production and improves their access to commercialization channels (ibid). Moreover, single women or women living in matrilocal and matrifocal societies, such as Windward Islands, also seem to have better possibilities of participation (Pollack, 2006; Smith, 2010).

Finally, women-only POs may provide a more protected environment that enhances female participation, as women may feel that their interests are better served, in comparison with male-dominated POs (Terstappen, 2010; Bergeron, 2010; Sen, 2009). Nevertheless, concerns are raised regarding the implication of dealing with men and women separately, where a lack of community support to the group coupled with limited economic profits can undermine performance and ultimately female participation (Pollack, 2006; Arce, 2009).

Certification interventions and their implementation

Training, new farming practices and PO support

Training, and broadly interventions for capacity building (for POs, producers, workers or plantations), are important interventions in many CS as they are designed to help participants meet the demanding standards required to obtain the certification. The adequacy of training, new farming practices – often referred to by CS as good agricultural practices (GAP) – and PO support is a key assumption for the CS to lead to improved farm outcomes, such as better farming practices for higher quality and yields, and efficient farm and producer group management. The following section presents synthesised findings related to these issues.

Training (n=30)

Adequate training, tailored to the producers' characteristics and needs, as well as sufficient and skilled extension staff able to properly cover the area and engage with producers, are key in successfully improving knowledge, adoption and implementation of GAP (Waarts et al, 2016; Aidenvironment, 2016; Stathers and Gathuthi, 2013; Donovan and Poole, 2014; Setrini, 2011; Barham et al, 2012; Jena et al, 2012), and can even lead to increased producer motivation and satisfaction (Waarts et al, 2014; Aidenvironment , 2016). Moreover, approaching training as 'development work' and not with a narrow focus on inspection and technical support can be more efficient in strengthening the institutional and financial capacity of POs, as Setrini (2011:309) reports. Tailoring training to overcome participants' limited literacy (CESU 2012) as well as gender constraints (Ellery, 2010; Pollack, 2006; Stathers and Gathuthi, 2013) are also reported to be important. Nevertheless, even

[85] This can happen pro-actively or after having their farms temporarily decertified due to absence, Lyon et al. (2010) report.

when training is successful, the capacity of the market to remunerate producers' effort to apply new practices can determine adoption or not (Aidenvironment, 2016).

Furthermore, financial constraints can hinder producers' participation, as producers may not be able to pay the training costs, or even afford to divert time from farming to training (CESU, 2012). In the case of wage workers, training is reported to lead to greater workers' participation in assemblies, higher self-esteem and even improved participation in community organizations outside the working place (Lyall, 2014). Compulsory training without remuneration, however, can be a problem, particularly if it does not result in higher wages (Bonnano and Cavalcanti, 2012).

Training related to CS can be provided by a variety of actors and appear to be highly context-specific. Training providers can vary from government and municipal agencies, and local and international NGOs to certification bodies and POs, including any combination of these actors. For instance, Utz is reported to add material on social and environmental practices to the content of already available training delivered by local extension services (Aidenvironment, 2016). In fact, training organised by local providers through institutionalised modalities, such as farmer field schools are reported to be more effective (Waarts et al, 2012 for RA training). In this multi-stakeholder context, the ability of POs to establish partnerships with private and public institutions and international buyers that can directly provide or finance extension support, as well as their capacity to re-invest in extension support is important (Abarca-Orozco, 2015; Setrini, 2011). POs with very large membership, may not be able to provide adequate (or any type of) training to their members, as Nelson et al (2013) report for the FT certified Ghanaian Kuapa Kokoo. 'Lead farmers' and local 'promoters' are reported to be important actors in successfully disseminating training material (Dowdall, 2012), however, may need close supervision in order to fulfil their role (Waarts et al, 2014). Finally, the effectiveness of training interventions may be hindered by lack of coordination between training providers (but also different CS), as cases of recommended practices contradicting past instructions and causing confusion, particularly between different CS, are reported by Lyon (2006).

Finally, it appears that training in CS standards and mechanisms have been less efficient than those in GAPs, as evidence from included studies point to a general lack of producers' knowledge and understanding of how CS operate (Shreck, 2002; Sutton, 2014; Staib, 2012; Heller, 2010; Makita, 2012; Nelson et al, 2013; Setrini, 2011; Trauger, 2014; Loconto and Simbua, 2010; Amekawa, 2001; Stathers and Gathuthi, 2013). Producers holding leadership positions in their POs (Walsh, 2004; Getz and Schreck, 2006) or in plantation workers' committees (Trauger, 2014), are reported to have a greater knowledge of CS and their mechanisms, suggesting that training regarding certification issues has been less successful in reaching the mass of certified producers, but only the more active minority, and that the expected diffusion the wider membership has not been effective. Overall this constitutes an important barrier because the lack of understanding of CS standards can be a major obstacle to standards compliance and uptake (Amekawa, 2001), hinder producers' control over the PO management and enable misuse of funds (Beuchelt, 2009), lead

to confusion and mistrust over the calculation and distribution of premia (Staib, 2012; Stathers and Gathuthi, 2013), and limit workers' capacity in addressing problems related to working rights (Lyall, 2014; van Rijn et al, 2016).

New farming practices / GAP (n=11)

New practices are a key certification input for improving yields and product quality, as well as ensuring that agricultural practices do not undermine the environment, but rather contribute to its protection. Successful adoption of such practices is therefore a crucial step in the impact pathway. Included studies highlight that while some recommended practices may be already part of producers habits and therefore easily adopted, such as the preparation and application of organic fertilizers for the 'organic by default' farmers (Terstappen, 2010; Dowdall, 2012), other practices may meet resistance, particularly if their application is costly and/or does not provide results in the short term (Amekawa, 2001).

The main barrier to adoption, however, appears to be the cost of applying new practices. Certification-recommended practices, such as renewing plants or harvesting in shorter intervals, may improve yields and quality in the long term, but in the short term they require extra labour, time and/or financial resources (Arce, 2009; Dowdall, 2012; Lyon, 2005; Donovan and Poole, 2014; Roy and Thorat, 2008a; Cofre et al, 2012). Furthermore, uptake can be low for practices that clash with the household economy, such as eliminating plants that farmers consider important for their income (Arce, 2009; Lyon, 2005), or simply shifting labour and financial resources from food crops, or other livelihood activities, to the certified crops (Naylor, 2014; Donovan and Poole, 2014b; Dowdall, 2012). Finally, certain farm management practices, such as record keeping, can be bothersome and time consuming for farmers lacking the necessary literacy skills and discourage adoption of standards (Lyon, 2005; Gomez-Cardona, 2012; Donovan and Poole, 2014).

Support to strengthen POs (n=42)

Certified POs tend to receive financial and technical support from a wide variety of actors, such as national and international NGOs, faith-based organisations, national and foreign governments as well as certification bodies and buyers (Bakker, 2014, Herman, 2010; Stathers and Gathuthi, 2013; Arce, 2009; Milford, 2004; Ellery, 2010; Dowdall, 2012; Lyon, 2005; Pongratz-Chander, 2007; Quaedvlieg et al, 2014; Smith, 2007; Abarca-Orozco, 2015; Sen, 2009; Shreck, 2002). Although this flow of aid occurs within the certification framework, is not always a direct input of certification. The phenomenon of certified POs becoming the focus of aid, is described as the 'honey pot effect' (Nelson et al, 2002), while the complexity of disentangling the effects of a single certification in a multi-donor and multi-certification context is highlighted by Nelson et al (2013) and Valkila and Nygren (2008).

Access to support appears quite important as a facilitator of adoption, given that often POs are too 'resource poor' to take the initiative of adopting a certification with all the associated demands. Therefore, several studies emphasise the critical role of external support to initiate and maintain certification-related projects and POs, a fact

which raises questions about the sustainability of certification in the absence of systematic external support (Bakker 2014, Herman, 2010; Stathers and Gathuthi, 2013; Arce, 2009; Milford, 2004; Ellery, 2010; Dowdall, 2012; Lyon, 2005; Pongratz-Chander, 2007; Quaedvlieg et al, 2014; Smith, 2007; Abarca-Orozco, 2015; Sen, 2009; Ronchi, 2002; Sutton, 2014; Nelson et al, 2013; Fairtrade, 2013; Roy and Thorat, 2008; Romanoff, 2010; Ouma, 2010; Fairtrade, 2013). There are instances in which external support is also provided by national governments through direct funding (Sutton, 2014), or the creation and support of consortia of certified POs (Abarca-Orozco, 2015; Pongratz-Chander, 2007). Other forms of external aid are debt settlements (Staib, 2012), extension support in kind (Vasquez-Leon, 2010; Walsh, 2004; Heller, 2010; Jaffee, 2006; Laroche et al, 2012; Asfaw et al, 2009b), establishing access to certified markets (Abarca-Orozco, 2015; Pongratz-Chander, 2007; Dowdall, 2012; Beall 2012) as well as income diversification programmes (Milford, 2014). Finally, certified POs and plantations are also reported to receive funds for non-productive programmes (Wilson, 2010; Makita, 2012; Ellery, 2010) and to mediate in the delivery of government and international aid services, as in the case of natural disasters (Mendez, 2002; Dowdall, 2012; Fraser et al, 2014).

Nevertheless, continued and extended external support, involves the risk of creating PO ‘aid dependence’, which can result in underdeveloped capacities and distorted budgets, thereby undermining their long-term sustainability (Donovan and Poole, 2014; Quaedvlieg et al, 2014; Riisgaard et al, 2009; Bacon, 2005; Shreck, 2002). This is even more important in the case of external aid provided by private companies, where support may be conditional on, or influenced by, the company's interests over those of the PO and the producers (Riisgaard et al, 2009). Other pernicious unintended effects of external support delivered through top down and paternalistic mechanisms include the risk of leading to producers' passivity (Shreck, 2002) or discontent (Dolan, 2010).

An important implication of the capacity of certified POs to broker government and donor funds is that it attracts producers, regardless of their expectations of benefits from certified markets (Jaffee, 2006; Milford, 2004; Milford, 2014; Pollack, 2006; Fraser et al 2014). In fact, external incentives may be the main, if not the only reason for adoption, suggesting that certification promises of benefits alone are not enough to incentivise adoption, as Beall (2012) and Kariuki (2014) report for RSPO and GlobalGAP Another matter is whether this support benefits a wide group of producers or is also subject to elite capture, an issue explored under the sections on distributional dynamics and contextual factors affecting PO’s effectiveness.

Pre-payment & credit (n=23)

Pre-payment can be an important mechanism that shapes adoption and outcomes, both for POs and producers, as it can provide POs with the means to buy the certified commodity before harvest, and therefore ensure the volumes needed to fulfil their contracts, while it provides producers with timely cash, enabling them to deal with the harvest costs, enhancing input use and minimising side selling (see more on side-selling in Section 4.5.3.1 - POs context: management, relationship with producer and with buyers). However, few examples of successful pre-payment implementation

are reported, both at the PO (Mendez, 2002) and producer level (Riisgaard et al, 2009), while cases where pre-payment is not the common practice (Milford, 2014; Walsh, 2004; Valkila and Nygren, 2008), or where pre-payment procedures established by the conventional market apply (Staricco and Ponte, 2015), are more commonly reported. Walsh (2004) suggests a barrier is the fact that CS lack the ability to force buyers to prepay, and buyers are reluctant to prepay due to experiences with POs failing to honour their supply contracts, without returning received pre-payments. That is another instance of principal-agent problems that CS are often unable to alter. Whatever the reasons, lack of pre-payment can have severe effects on POs and producers, as it can cause shortages of working capital and undermine the ability of POs to provide loans to their members during harvest (Valkila and Nygren, 2008) but also lead to payment delays (Milford, 2014; see section on ‘Payment delays’).

Access to credit is often reported to be the prime incentive of producers for joining a PO, whether certified or not (Beuchelt, 2009; Lyon, 2007a). Nevertheless, certified POs appear to have better access to credit than non-certified ones, through specialised international financial institutions which accept export contracts as collateral (Laroche et al, 2012; Walsh, 2004). Certified POs are also reported to receive long-term loans from aid agencies, and to acquire bank loans with favourable terms due to secured access to international markets and transparent fiscal policies adopted as a result of CS standards (Lyon, 2007a; Lyon, 2007b). Finally, Fairtrade premia are also reported to be used to create reserves which can be accessed as credit to be used as working capital (Ronchi, 2002).

As a result, certified POs are reported to offer credit to their members on better terms than conventional traders, especially compared to private money lenders (Milford, 2004; Nelson et al, 2002; Donovan and Poole, 2014; Dowdall, 2012); the only reported exception being a coffee PO in Nicaragua which apparently charged up to twice the interest demanded by private export companies (Valkila, 2009). Certified tea plantations in India are also reported to provide better access to credit to their employees (Stathers and Gathuthi, 2013), as well as letters of support for workers seeking access to formal credit (Nelson and Martin 2013).

Nevertheless, some interesting dynamics related to credit are reported by the included studies. First, although credit tends to be on better terms – or even the only terms accessible to small producers – the cost is reported to still be high compared to producers' gains from certification (Dowdall, 2012), while the amounts offered appear insufficient to cover production costs, and much less investments, while usually long-term credit is not available (Donovan and Poole, 2014b; Beuchelt, 2009;Valkila and Nygren, 2008). Since these loans are insufficient to cover production costs, producers are reported to use the funds to cover living costs rather than re-invest them in the farm, creating a vicious circle of low yields, net losses and high indebtedness (Beuchelt, 2009; Wilson, 2010; Lyon, 2007b). As an exception, Dowdall (2012) reports on Fairtrade producers using PO credits to repay loans acquired to buy land.

Second, side-selling may be an unintended outcome of the failure to provide pre-payment or credit. Thus, as a result of the limited capacity of certified POs to provide sufficient and timely credit and/or pre-payment to their members, certified producers in need of larger, earlier and/or emergency credit, take loans from private merchants, neighbours or employers in exchange of future harvest, and therefore end up selling their product at lower prices outside of certified channels (Milford, 2014; Cramer et al, 2014b; Jena et al, 2012). This phenomena, known as 'crop- mortgaging' (Wilson, 2010), can be accentuated in periods of price or harvest crises and can lead to high levels of indebtedness, particularly if credit surpasses the value of the final product harvested (Beuchelt, 2009)

Third, it is reported that pre-payment and credit provided by certified POs is often used to cover the increased labour and input costs resulting from certification, meaning that the extra liquidity is consumed by the higher costs of certified production (Jaffee, 2006). If coupled with delayed final payments, a common practice in certified POs as discussed in the section on 'Payment delays', and high interest rates, then producers may enter a vicious cycle of debt (Dowdall, 2012).

Fourth, better-off producers appear to get better access to credit. For example, POs offering better credit schemes may also have more demanding quality criteria, as in the case of a Nicaraguan coffee PO, where better access to credit was in practice only available to farmers who were able to satisfy higher quality standards (Donovan and Poole, 2014). Similarly, land owners who could present land titles as collateral are also reported to have preferential access to PO credit (Wilson, 2010).

Fifth, CS-related credit gains significance in cases where the state does not provide subsidies to farmers, or when local credit providers cease to support small producers, as during the 2000-2004 coffee crisis in Nicaragua, for instance (Valkila and Nygren, 2008; Wilson, 2010). Thus the role CS-related credit plays in filling an important gap, can in some cases be a determinant of more effective engagement of POs and producers.

Minimum Price (n=18)

A key intervention specific to the Fairtrade certification is the guarantee of a minimum (or floor) price, which aims at providing producers with stable incomes even during periods of price crisis (Fairtrade, 2016). This kind of intervention was one of the cornerstones of Fairtrade's ToC, as a key aim was (and is) to protect producers from the vagaries of increasingly liberalized and volatile agricultural commodity markets. This sort of price insurance is probably one of the interventions that certified producers are most aware of in Fairtrade POs. The synthesis of findings identifies the following main points.

First, the minimum price indeed acts as a safety mechanism during price slumps and enables POs to maintain their market share by offering higher prices to their members (Ronchi, 2002; Pongratz-Chander, 2007; Valkila and Nygren, 2008; Abarca-Orozco, 2015; Milford, 2014). It is also reported to provide stability and allow POs to make long-term investments, while protecting them from price fluctuations

(Ronchi, 2002; Pongratz-Chander, 2007). The mechanism is obviously effective when there is compliance by buyers, who pay the minimum price, but this is not always the case (Smith, 2010; Vasquez-Leon, 2010; Staricco and Ponte, 2015).

In periods when conventional prices rise above the Fairtrade minimum price, however, the only difference between the conventional and the Fairtrade market is the Fairtrade premium and the floor price becomes irrelevant (Milford, 2014; Nelson and Martin, 2013; Moore 2010), while producers lose incentives to invest extra labour in certified production and look for the best opportunities outside of certified selling channels (Arce, 2009). Furthermore, for some commodities, like cocoa (Ryan, 2001) or tea (Nelson and Martin, 2013), this mechanism has never been activated, confirming the ToC assumption that market volatility needs to be a problem for CS price interventions to have an impact. Coffee is probably the certified commodity for which the minimum price is the most significant intervention.

Second, besides market fluctuations, the portion of production sold as Fairtrade also matters, as limited amount of product sold to the certified markets results in limited protection from the minimum price mechanism. For POs that only sell a small percentage of their product to the Fairtrade market (see section on 'Markets'), the benefit of the minimum price vanishes as overall prices end up mimicking the fluctuation of the conventional prices (Walsh, 2004; Smith, 2007).

Third, the Fairtrade minimum price, even when it is above the conventional market, is often reported to be insufficient to cover production costs (Smith, 2010; Walsh, 2004; Dowdall, 2012; Terstappen, 2010; Jaffee, 2006). Inflation affects the real value of the minimum price, and criticisms are expressed towards Fairtrade for failing to adjust prices (Dowdall, 2012; Staricco and Ponte; 2015). **Figure 4** illustrates the long periods during which the minimum price was maintained at the same nominal level. An exception is reported by Shreck (2002), regarding bananas from the Dominican Republic, where FLO has shown willingness to improve and adapt its price regulation to the specific context.

Price premium

This section synthesises findings related to the implementation dynamics of CS price premia. We find that the literature does not always differentiate between different uses of price premia, for instance when the price differential goes directly to the producer, often in the form of second payments (i.e. Jaffe, 2006), compared with premia that are supposed to be managed collectively and invested in projects that benefit the wider community. This is what is normally called 'social' or 'community' premium, though the main CS applying this modality of intervention, Fairtrade, only refers to the 'premium'.[86] Utz applies a premium, usually recommended for quality and achievement of social standards but leaves it at the discretion of the buyer. For many studies the line between the two is not always clear. This section will attempt to

[86] http://www.fairtrade.net/about-fairtrade/what-is-fairtrade.html

provide a separate discussion as the dynamics, barriers and facilitators are specific to each type of use of the price premium.

Price premium as additional payment (n=22)

The main finding regarding certification price premia (understood as an additional payment per unit of output sold to certified markets and typically paid at the end of the season/year) is the often reported insignificance of such payments at the producers' level. A series of reasons appear to lead to this. First, premia are subject to important deductions from the total gross price received by the PO prior to the final payment to the producer. These can be due to tax payments (CESU, 2012), PO administrative costs (Milford, 2014; Valkila, 2009; Dowdall, 2012), or producers' debt cancellation (Dowdall, 2012; Walsh, 2004; Wilson, 2010). Premia are also reportedly used to cover certification costs (Rueda and Lambin, 2013; Dowdall, 2012), increased production costs (Cepeda et al, 2013), and product losses due to CS related stricter quality standards and restrictions in chemical use (Smith, 2010), in other words, to finance the certification process.

Second, for the POs that only sell a limited proportion of their product to the certified market, but distribute the premia to a large base of members, the premium benefits may get so diluted that they become quantitatively insignificant for the producers (Heller, 2010; Shreck, 2002; Nelson et al, 2013; Nelson and Martin, 2013). This appears to be an internal contradiction of the CS market, as the large size of POs is actually necessary to gain economies of scale (Milford, 2014) and compete in the certified market (Ryan, 2011). As a result, the price differential between certified and noncertified markets may become very small and producers lose the incentive to invest in extra labour required for certified production (Beall, 2012; Loconto and Simbua, 2010; Milford, 2014; Valkila, 2009; Laroche et al, 2012; Jaffee, 2006). In fact, what appears to be significant to producers is not the general premium, but the additional premium for organic certification (Jaffe, 2006; Laroche et al, 2012; Valkila, 2009; Walsh, 2004).

Third, the unstable nature of the Fairtrade price premium can also lead to higher levels of PO indebtedness. According to some studies, POs apply for larger loans based on their expectations about the premium and are forced to default if they are unable to recover the expected Fairtrade price, in which case losses are likely to be deducted from the final price paid to farmers (Abarca-Orozco, 2015; Beuchelt, 2009; Fraser et al, 2014). Another issue reported to affect the significance of the Fairtrade price premium is non-compliance of buyers regarding Fairtrade premium payments and the limited power of FLO to enforce the standard, due to fears of losing an important market for producers (Shreck, 2002). Finally, the premium does not operate as an incentive to improve quality since it is provided to the PO on the basis of the quantity sold as certified, irrespective of quality, thereby creating a disadvantage for plantations growing produce such as Darjeeling tea, which tend to sacrifice quantity to prioritise quality (Moore, 2010).

Premium for 'social' or 'community' investments

POs and producers (n=21)

Findings on implementation dynamics regarding the use of the 'social' premium reveal difficulties and conflicts related to its management. First, some higher-level producer unions appear to require written projects proposals by the local POs in order to release the 'social premium' funds. This, however, can be a challenging task for producers with low literacy and little education and actually result in funds being blocked at the PO union level, while POs struggle to produce a proposal (CESU, 2012; Jena et al, 2012). Second, local group dynamics and power relations clearly affect the way the 'social premium' can be 'collectively' managed to benefit the wider community (Naylor, 2014; Staib, 2012; Sen, 2009). For example, Naylor (2014:142) reports that in politically heterogeneous and polarised communities in Chiapas, Mexico, the realisation of such community development projects can be very challenging, a finding which strongly contrasts with the Fairtrade view of 'producer communities as homogenous and economic development plans as universally beneficial'. Third, but closely related to the previous point, cases of elite capture in the decision making of the 'social premium' are also reported (Sutton, 2014; Cramer et al, 2014b; Staib, 2012; Dolan, 2010). For example, Cramer et al (2014b) report that social premia were invested on projects that disproportionately benefited wealthier members of the organisations, such as an electricity line that only connected the houses of a few of the wealthiest members, or a fee-paying school in areas of widespread poverty.

Furthermore, the assumption that the premium is usually invested in social infrastructure to the benefit of the whole community is questioned by several studies, which show that the uses of the premium include all sorts of outlays and objectives such as paying for certification costs to gain independence from donors (Quaedvlieg et al, 2014; Valkila, 2009; Valkila and Nygren, 2008); investing in infrastructure and credit services that can attract new members; or simply covering running costs (Setrini, 2011; Terstappen, 2010; Walsh, 2004; Smith, 2010; Cepeda et al, 2013; CESU, 2012; Shreck, 2002). Producers are also reported to collectively decide to use of social premia to pay off PO debts, often resulting from previous mismanagement and corruption (Fraser et al, 2014; Mendez, 2002; Minten et al, 2015). Additionally, in times when the conventional price is approaching the certified price, POs are reported to divert the 'social' premium to boost their price and maintain their market share (Dowdall, 2012; CESU, 2012; Smith, 2010; Sutton, 2014). Although such uses of 'social premium' can benefit producers by building stronger, healthier and more independent POs, it is also highlighted that keeping the social premia entirely or primarily in the administrative budget of the PO does not benefit the wider community as it is supposed to do (Cepeda et al, 2013), and may not even benefit the producers if the management of the PO is corrupted or inefficient (Staib, 2012).

Plantations (n=11)

Studies focusing on plantations, reveal slightly different implementation dynamics regarding the ‘social premium’ compared to POs. A controversial point appears to be the use of the premium as a cash bonus. This is not allowed by FLO (Smith, 2010; CESU, 2012; Brown, 2012), however, in many cases workers are reported to prefer to have it distributed as a wage supplement instead of investing it in community projects, particularly in cases where the salaries are very low and do not cover workers’ living costs (Brown, 2012; Smith, 2010; Staricco and Ponte, 2015), or when many workers are migrants who do not have links with the local communities (CESU, 2012; Smith, 2010; Trauger, 2014). For example, Staricco and Ponte (2015) report how workers in Argentinian wineries employed both agency and creativity to overcome FLO restrictions on dividing the premium among workers in cash or kind, by creating a basket of ‘healthy products’ that was distributed to each worker, framed within a wider educational/health project. A case of premium being used as a wage subsidy is also reported (CESU, 2012).

A second controversial point is that of plantation management control over and manipulation of the ‘social premium’. Since the ‘social premium’ is for the workers, it appears that plantations have little financial benefit from certification, and therefore no incentive to expand their certified market share (Setrini, 2011) or to maintain certification, besides satisfying their exporters and their workers (Brown, 2012). Nevertheless, the synthesis of findings reveals that plantations’ managements far too often make substantial decisions over the use of the premium on behalf of the workers, usually by controlling and manipulating the committees responsible for managing the premium funds (Brown, 2012; Besky, 2014; Makita, 2012; CESU, 2012; Moore, 2010). A characteristic example is provided by Moore (2010) who describes how cows purchased with the Fairtrade premium to provide tea workers' communities in India with milk and extra income from milk sales, were actually part of the management’s strategy for getting free manure, while workers perceived the work involved in taking care of the cows as a burden due to the unsatisfactory local milk market.

Such attempts at manipulation, but also power imbalances between workers and management, can lead to workers mistrusting, rejecting, misusing or even destroying projects funded by certification premia (Setrini, 2011; Moore, 2010). On the contrary, reported cases of successful implementation of premia, (e.g. computer centres, scholarships, energy, credit and community tourism programs), are attributed to the existence of more democratic structures, less hierarchical management and better collaboration between workers and management (Moore, 2010). A small plantation size that allows for greater individual worker participation, and certification standards tailored to the specific plantation and its context are also reported to contribute to more successful premium investments (Moore, 2010).

A third point of controversy, closely related to the previous one, is that certification premia (along with other certification benefits) are reported to be used in plantation contexts as a reward for ‘well-behaved’ workers, but also as a mechanism to increase workers’ compliance with higher production expectations (Setrini, 2011; Makita, 2012; Lyall, 2014; Schelly, 2011; Brown, 2012). Maquita (2012:100), for

instance, highlights that the invisibility of Fairtrade in the plantation leads workers to assume the premium to be the result of the generosity of the management, increasing therefore the 'degree of patron compliance' among workers.

Many of these instances of the use of premium for purposes other than the expected, given recommendations from the CS, illustrate that the notions of 'community' and of homogeneous groups of producers or workers are problematic. In the end, the uses of premia will depend on the particular configuration of interests within a PO or plantation, and on the power of some to bend rules and meet the expectations of some of the concerned constituencies. As is argued in Section 1, these interventions do not happen in a social, economic and institutional vacuum.

Payment delays (n=16)

Payment delays, sometimes up to several months after delivery, are one of the main complaints of producers and also a main reason for leaving (or never joining) a certified PO, or for side-selling (Dowdall, 2012; Bacon, 2005, Fayet and Vermeulen, 2014; Jaffee, 2006; Mendez, 2002; Milford, 2004; Milford, 2014; Rotter, 1999; Sutton, 2014; Valkila, 2009; Valkila and Nygren, 2008; Bagama et al, 2014). It is not always clear, however, if these delays originated at the PO, or the buyer/market level, or somewhere in between, and whether or how they are related to the certification framework. Whatever the origin, this is an implementation failure that can cause a certified PO to fail to meet its objectives and retain membership. The problem is also related to difficulties in arranging credit or pre-payment, as show in section above (Mendez, 2002; Milford, 2004). Staricco and Ponte (2015) situate the origin of the problem between the buyer and the PO level, attributing payment delays to excessive tolerance of Fairtrade International towards buyers and lack of internal controls at the PO level. Sutton (2014), on the other hand, suggests that payment delays in the case of Fairtrade coffee POs in Tanzania were a combined result of buyers' delays and Robusta auction delays, but also PO mismanagement.

Payment delays cause farmers, particularly those with less resources, no alternative sources of income, or in need of quick cash, to prefer the lower prices of the intermediaries who pay directly and on the spot (Bacon, 2005, Dowdall, 2012; Fayet and Vermeulen, 2014; Jaffee, 2006; Rotter, 1999). However, cases of intermediaries paying the same as (Milford, 2004; Valkila and Nygren, 2008), or even better than certified POs (Donovan and Poole, 2014, 2014b) are also reported. Payment delays also appear to create mistrust among members who fear they will never recover their payments (Smith, 2007), and there are cases of POs failing to make the final payments (Milford, 2004; Milford, 2014). A possible advantage of not paying in full at harvest is, as Jaffee (2006) argues, that payments are spread through the year, which means that certified producers can receive cash just before the harvest (i.e. a period entailing extra labour costs) when non-certified producers have already run out of cash and start borrowing. Nonetheless, as shown above, most studies tend to focus on payment delays as a problem and barrier rather than a positive aspect.

Labour standards (n=18)

Findings related to certification labour standards point to the importance of the legislative context and the degree to which this is enforced. If national labour laws cover and exceed the guarantees offered by CS, and are properly enforced, benefits from standards become obsolete (Moore, 2010; Valkila and Nygren, 2008; Staricco and Ponte, 2015; Nelson and Martin, 2013; Raynolds, 2012). A characteristic case is that of Indian plantations and the Plantation Labour Act (PLA). Although the proper enforcement and effectiveness of the PLA have been widely questioned and debated (i.e. Bhowmik, 1992; Saikia, 2008; John and Pallavi, 2013), Moore (2010, p.21) argues that it still guarantees rights 'above and beyond those guaranteed by [Fairtrade]'. Nelson and Martin (2013, p.128) also comment on the co-existence of certification and the PLA, reporting that although there were improvements in services after certification, these did not lead to 'a radical transformation as the Plantation Labour Act already meant workers had a certain level of service provision'. Commenting on the case of Nicaraguan labourers working in coffee farms, Valkila and Nygren (2008, p. 331) also question whether 'Fairtrade really raises the bar on labour standards above the existing norms, or whether Fairtrade just complies with the labour standards already recognized in the Nicaraguan legislation'. On the other hand, if national legislation is less demanding than the certification standards, or subject to frequent violations, then workers are reported to see significant improvements, such as paid annual and maternity leave (Raynolds, 2012, 2014; Schelly, 2011; Smith, 2010). For example, Fairtrade is reported to go beyond labour laws in Ecuador, while also acting as a law enforcement mechanism for certified plantations (Raynolds, 2012, 2014; Schelly, 2011). Nevertheless, some studies report that in the absence of strong supportive legal frameworks, poor labour conditions persist if standards are limited to recommendations that are not binding for the employer or simply not properly enforced (CESU, 2012; Bonanno and Cavalcanti, 2012; Setrini, 2011). For instance, while some authors report that workers in certified plantation benefit from reduced exposure to chemicals and adequate protective equipment (Brown, 2012; Raynolds, 2012), others report inadequate uniforms and pressure to re-enter fumigated spaces (Schelly, 2011; Trauger, 2014).

Moreover, law enforcement and standards application may be inadequate for workers and plantations in some contexts (Raynolds, 2012, 2014; Stathers and Gathuthi, 2013; Staricco and Ponte, 2015; Nelson and Martin, 2013; Smith, 2010; Moore, 2010). The most controversial issue is overtime restrictions and the resulting inability of the regulation to accommodate workers who would prefer to work overtime, as in cases where minimum wages are below living costs meaning that such restrictions may lead to a decrease in income below living standards (Raynolds, 2012, 2014; Stathers and Gathuthi, 2013; Staricco and Ponte, 2015; Nelson and Martin, 2013). Nelson and Martin (2013) report that over-time restrictions not only significantly decreased daily wages, but also undermined the capacity of workers to repay loans taken on the basis of their wage including over-time, while cases of workers leaving their jobs due to this restriction are also reported. Over-time restrictions were not welcome by management neither, as they supposedly complicated the logistics of rotating workers according to maximum hours permitted

(Staricco and Ponte, 2015) and affected the plantations' competitiveness during peak periods (Raynolds, 2012).

Overall, it appears that CS do not provide extra benefits when labour laws exceeds CS standards, but are also limited when the labour market and laws do not provide enough by themselves. A key example is the minimum wage. Many studies report that the minimum wage for workers, both in plantations and on smallholders' farms, is insufficient. Regarding certified plantations, minimum wage requirements tend to align with nationally stipulated minimum wages, which are usually below living costs and insufficient to support a family (Raynolds, 2014; Schelly, 2011; Smith, 2010). Additionally, Heller (2010), reporting on wage workers working in certified coffee farms in Guatemala, describes how difficult is to enforce this standard in practice, due to a lack of precision in the determination of the minimum daily wage, as well as small producers' financial limitations, which make it difficult for them to provide the minimum wage to their workers. Smith (2010) also underlines that guaranteeing minimum wage payments brings limited benefits if minimum wages stipulated by national laws do not cover living costs.

Finally, while insecurity in the labour market is reported to increase the value of certification as a source of legitimate employment (Brown, 2012), it also appears to undermine CS capacity to empower workers and enforce their ability to defend their rights, as workers in contexts of labour insecurity are often reported to be afraid to confront their employer (even in cases of payment delays), despite certification status (Lyall, 2014; Schelly, 2011; Setrini, 2011; Sen, 2009).

Overall, several studies question the relevance of CS labour standards and place emphasis on the role of national labour institutions and companies as the key determinants of labour outcomes (Nelson and Martin, 2013; Staricco and Ponte, 2015; Cramer et al, 2014a). For instance, Nelson and Martin (2013), report that employment formalisation in tea plantations was driven by a combination of labour shortages and legislative obligations, which resulted in new recruits being rapidly moved into permanent contracts, and not by the presence of CS. Similarly, Cramer et al (2014a) find that in tea plantations in Uganda, the best labour standards were found in large-scale, multinational corporations where certification was not present and was unnecessary from the employer's point of view.

Costs of certified production (n=37)

A key assumption necessary for the CS to lead to increased incomes and profits is that certification benefits exceed the costs. The implementation dynamics of CS are certainly influenced by the level and different types of costs associated with certification, whether direct or indirect, linked to the investments to meet standards or to the actual costs of registration. This section synthesises findings on the different costs involved in certified production, including organic certification.[87]

[87] Organic certification is commonly combined with social standards, particularly with Fairtrade. It is very difficult to disentangle the costs of organic from other certification in these cases, as producers, cooperatives and researchers tend to treat organic and social standards

A majority of studies focus on a key indirect cost of certification: workload. An overall increase in workload (i.e. more frequent harvesting and more laborious post-harvest processing) is widely reported to be required to meet increasingly strict and demanding - both in terms of quality and compliance - certification standards (Gomez-Cardona, 2012; Shreck, 2002; Setrini 2011; Valkila and Nygren, 2008; Smith, 2007; Laroche et al, 2012; Jaffee, 2006; Moberg, 2008; Setrini, 2011). Workload seems to be particularly affected in those CS which also include organic standards, as the combination of organic-social standards tends to increase not only the use of hired labour, but also the workload of all involved in production, including family labour (Jaffee, 2006; Setrini, 2011). The wage bill may increase because of additional hours of hired labour that need to be purchased, perhaps at a higher price if labour standards are enforced (Trauger, 2014), which can be positive as a demand injection for the local economy, but the workload can also be assumed by the wage workers without the necessary compensation for the extra work (Setrini, 2011). A coping mechanism for poorer producers in such situations is to draw more on cheaper labour sources, such as household members or (illegal) migrant labourers (Shreck, 2002; Trauger, 2014). This can at times lead to more reliance on child labour, which runs against one of the basic labour standards supposedly enforced by most CS (Milford, 2014; Dragusanu and Nunn, 2014; Valkila and Nygren, 2008), despite the fact that labour standards regarding child labour are ‘painstakingly explained’ to producers and are being thoroughly applied, Naylor (2014, p. 137) points out.

The increase in labour costs associated with the implementation of standards is gendered in a variety of ways. Post-harvest tasks aiming at higher quality occur in proximity to the producers’ houses and therefore are mainly undertaken by women (Jaffee, 2006; Nelson et al, 2002), although Lyon et al (2010) note that certification may provide the resources to mechanise these steps and therefore reduce women's workloads. Plot maintenance tasks, on the other hand, are considered mainly male work, as Jaffee (2006) reports for certified coffee production in Mexico.

Another source of extra workload, for producers, POs and plantations, is the amount of paperwork required to meet organic and social standards, including daily record keeping or internal monitoring activities (Brown 2012; Dowdall, 2012; Loconto and Simbua, 2010; Walsh, 2004; Trauger, 2014; Beuchelt, 2009; Naylor, 2014; Balineau, 2011; Raynolds, 2012; Abarca-Orozco, 2015; Valkila and Nygren, 2008; Jari, 2013). The time and effort required to attend meetings and training sessions is another factor that increases certified producers’ workload, which in combination with extra paperwork duties and increased labour, may come at the expense of other farming or non-farming activities (Gomez-Cardona, 2012; Milford 2004, 2014; Naylor 2014; Pongratz-Chander, 2007; Setrini, 2011; Walsh, 2004; Jaffee, 2006).

Moreover, stricter quality criteria as a result of certification, as well as transition from conventional to organic production, can cause product losses and therefore have an

as one compact certification/intervention. Therefore, costs related to organic production are presented here as part of the certification costs.

indirect effect on production costs (Nelson and Martin, 2013; Shreck, 2002; Abarca-Orozco, 2015). Certification-related investments in infrastructure, inputs, equipment, as well as traceability systems are an additional considerable source of cost augmentation for POs and plantations alike (Asfaw et al, 2009b; Moyo and Mugabe, 2014; Schelly, 2011; TWIN 2012).

There are other implementation and registration costs that may be an obstacle to adoption. First, entry fees required to sign up to certified POs, although usually low, can also be a barrier for poorer producers (Mendez, 2002; Milford, 2014; Walsh, 2004). Second, regulations related to attending meetings may include fines (or even exclusion), which adds pressure on poorer producers (Milford, 2014). These can lead some well-performing POs (i.e. no payment delays, attractive credit scheme) to have stricter membership regulations, which act as mechanism to control the size of the membership and result in excluding producers not able to comply (Milford, 2004). Third, direct costs of certification and inspection can be significant, as they are reported to put off some POs due to a lack of resources (Dowdall, 2012), or a lack of transparency on how these fees are used by the certifying bodies (Loconto and Simbua, 2010). However, the variation in direct certification costs across CS is huge. Ruben and Zuñiga (2011) report a range between 2 and 4 US$ cents/lb in Nicaragua for Fairtrade coffee production, which contrasts with very large sums incurred by large-scale fruit producers in Chile to get GlobalGAP certification, an average of over 4,000US$ per hectare annually (Cofre et al, 2012). Of course, the revenues per ha of large-scale fruit producers in Chile are likely to be much higher than those of smallholder coffee producers in Nicaragua.

Given this account of CS-related costs, it is not surprising that the ability to afford the increased costs involved in social and organic standards, mainly in terms of labour but not only, is reported to determine the producer decision to join certified POs (Milford, 2004, 2014; Jaffee, 2006; Lyon et al, 2010; Trauger, 2014; Waarts et al, 2012). POs are reported to depend on external aid (see Section 4.5.3.1), as well as on participation in unions of POs and consortia (Terstappen, 2010) in order to obtain and maintain certification, or to drop out until they are able to meet the costs (Sutton, 2014).

In the end the key question is whether the price incentives obtained through CS compensate for the additional workload and costs discussed in this section. With the exception of Stathers and Gathuthi (2013), most studies report that higher costs involved in certified production are not adequately compensated for in terms of price incentives (Smith, 2010; Laroche et al, 2012; Bacon, 2005; Beuchelt, 2009; Dowdall 2012; Jaffee, 2006; Lyon et al, 2010), suggesting that the core assumptions of the ToC, i.e. that premium and markets are sufficiently remunerative and that certification benefits exceed the costs do not always hold. A limitation of these studies is that they are based on basic comparisons of price premia and certification costs, without engaging in a systematic and more comprehensive cost-benefit analysis of the process, including intangible benefits such as knowledge or access to new market channels. Until more systematic cost-benefit comparisons are made it will be hard to give a more rigorous assessment on the extent to which benefits outweigh costs.

Nonetheless, the fact that a large number of studies raise this problem suggests the cost-benefit ratio does not always meet expectations.

Monitoring and auditing (n=12)

Effective monitoring and enforcement mechanisms are another set of basic assumption of the ToC for CS to bring real changes, given that each certification may carry multiple standards with a long list of requirements that need to be audited. Unfortunately, and despite the importance of this issue to understand the process of certification and the barriers and facilitators for its effectiveness, relatively few studies report relevant evidence.

Internal monitoring is a key element of all certified POs, who are required to create and maintain an internal control system to ensure compliance with standards and to facilitate the work of external auditors. Internal monitoring is also expected to safeguard the standards' integrity, for instance by keeping certified producers from speculating on price differentials by selling non-certified products produced by other producers as their own (Heller, 2010). Moreover, internal monitoring systems are opportunities to producers, particularly women, to become actively involved in their POs and learn extra skills they would not gain by selling to intermediaries, such as inspecting product quality (Lyon et al, 2010; Walsh, 2004). Nevertheless, internal controls can create tension within producer groups, as producers may be reluctant to monitor fellow producers, particularly when personal commitments to family and friends are at play (Lyon, 2005, 2006).

The standards reviewed in this report are all third-party certified, which means that there is an independent body that determines whether the final product complies with the specific standards. This is commonly done by external auditors who visit the POs, plantations and producers and perform random checks. Two main points are usually raised by reviewed studies regarding external audits. First is the lack of transparency and accountability of auditing bodies, as well as the lack of feedback given to the audited producers (Lyon, 2005; Setrini, 2011; Larsen et al, 2014). Moore (2010) particularly questions the integrity of FLO-Cert audits, the only body allowed to certify Fairtrade products, as its financial interests are exclusively linked with Fairtrade and this may cause reluctance to reveal serious problems in Fairtrade compliance that could damage the credibility of the brand, while Francesconi and Ruben (2014) also report that there is tension between the imperative of expanding membership in Fairtade certified POs and the capacity to monitor and audit POs with very large membership bases.

A second point raised is the degree to which external audits can be manipulated by POs and plantations to simulate compliance. For instance, Heller (2010) describes how producers make last moment changes in their production practices in order to pass the inspections. Regarding plantations, Moore (2010) and Schelly (2011) comment on how pre-announced yearly audits allow managers to get prepared by distributing new uniforms and protective material, as well as by selecting the most favourable workers to be interviewed, while workers avoid expressing any complaints out of fear of losing certification-related benefits.

Moreover, studies suggest that audits often lack the ability to thoroughly understand producers' reality and make appropriate recommendations, especially when they are conducted in just a few days for POs including thousands of members (Heller, 2010; Trauger, 2014; Sen, 2009). For example, Sen (2009) describes how language barriers, limited interaction with workers and poor understanding of the socio-cultural context lead to FLO auditors making inappropriate recommendations that actually could harm the interests of female farmers.

Spill-over and unintended CS effects (n=16)

This section synthesises reported findings on spill-over and unintended effects. The following points are identified.

First, CS appear to have an upward influence on local prices, as higher prices offered by certified POs are reported to intensify competition among local intermediaries, who in an attempt to compete end up offering higher prices as well, an effect described as the 'competitive yardstick effect' (Milford, 2014; Walsh, 2004; Laroche et al, 2012).[88] Ronchi (2002) reports that this effect also applies to hired labour wages by smallholder producers, however, it contradicts other studies reporting no upward effect on hired labour wages related to certification (Shreck, 2002; Valkila, 2009; Dragusanu and Nunn, 2014; Heller, 2010).

Second, spill-over effects are reported on the adoption of GAPs, as non-certified producers are reported to participate in GAP training (Nelson and Martin, 2013) and to adopt certain new practices (Quaedvlieg, 2014). Nevertheless, failure to spread the use of GAPs among non-certified producers, despite efforts of certified POs to improve local industry standards, are also reported (Nelson et al, 2002).

Third, the certification industry appears to create more employment opportunities not only for labourers from the community due to increased labour needs, particularly when it comes to organic production (Jaffee, 2006), but also to generate demand for 'skilled employment related to production standards and premium use', such as auditors, agricultural extension staff, or community development workers and administrators (Smith, 2010, p. 100). It is not clear, however, how these jobs are financed, although several studies suggest that producers, POs and plantations assume a large part of these costs, using CS premia but also external aid funds.

Fourth, CS appears to contribute in transforming the local economies. Sen (2009), for instance, reports how certified tea plantations in Darjeeling ended up incorporating the 'illegal' but organic by default production of smallholders in the area which was sold only locally, to offset the decrease in production due to restrictions in chemical use.

Fifth, certification related farm documentation is reported to be used to settle land disputes against uncertified smallholders, which do not have formal documentation to support their claims, as Kohne (2014) reports for the case of RSPO, or to create 'de facto property', as such documents assign control over land to the register owner and

[88] Upward influence of CS on local prices is also suggested by Fort and Ruben (2009), study included for RQ1.

can be used to register plots with state or private corporate entities (Lyon et al, 2010).

Multi-certification (n=8)

Multi-certification is an important fact to consider not only in terms of impact attribution, but also in terms of understanding the CS implementation dynamics, as the synergies developed between different standards can be influential. There is a tendency for POs and plantations to seek as many certifications as possible in an attempt to 'capture the widest market' and cases of POs holding two or more different certifications are commonly reported (Trauger, 2014, p. 1092). In fact, some standards are reported to have a significant overlap, and therefore once one certification is obtained, it is relatively easy to meet the demands of others, as Nelson and Martin (2013) report for Fairtrade and Rainforest Alliance in the case of Indian tea plantations.

The dual Fairtrade-organic is the most typically studied (this qualitative synthesis includes 39 studies on this combination), possibly because it is the most commonly found. Several drivers explain this combination. First, for commodities such as coffee, organic certification is reported to be a requirement for reaching Fairtrade markets, as demand for Fairtrade alone is very limited (Valkila, 2009). Strong demand for organic products also means that while producers are uncertain to receive the Fairtrade premium, organic markets are more likely to remunerate for the total of the harvest (Valkila, 2009). Additionally, organic certification offers the opportunity to sell more crops as organic once the farm is certified, and not only the certified commodity as in the case of social standards (Makita, 2011). On the other hand, Fairtrade certification is reported to be used to give producers some extra income during the conversion to organic, a period during which production is likely to decrease and producers do not yet receive the organic premium (Makita, 2011). Nevertheless, the coexistence of organic and other standards is not always the optimal option. Cepeda et al (2013), for instance, report on Ecuadorian cocoa POs dropping Rainforest Alliance in favour of maintaining organic certification, which was perceived as opening more markets and having a more assured demand. On the other hand, the costs of organic certification appear to be the main reason for which coffee producers do not join Fairtrade-organic POs (Milford, 2014; Jaffee, 2006; Abarca-Orozco, 2015), therefore suggesting that organic certification can be a barrier to Fairtrade adoption for those producers who cannot afford the extra costs of organic production.

Conclusion

The synthesis of findings on implementation dynamics has provided evidence on multiple instances in which implementation does not happen as it is supposed to and is instead influenced by a myriad of factors, often contributing to lower than expected effectiveness. Many studies tend to highlight the mismatch between the expectations raised and the claims made by many CS one the one hand and the participants' lived experiences on the other, particularly with regard to monetary benefits, but also working conditions and female participation. The configuration of factors that shape

the dynamics of implementation of CS is such that it is hard to see how CS can overcome or even mitigate deep-rooted problems and constraints in the socio-economic and political context in which they are implemented. The synthesis has given several examples of how adoption is influenced by the certification costs, and of how they compare to the actual benefits – a relation that is turn influenced by the presence of external support from buyers, NGOs or other agencies, and by the availability of pre-finance. Likewise, the effective implementation of premium usage and the applicability and suitability of labour standards also depend much on the local and national-level factors that affect POs', producers' and plantations' decisions, in particular the nature of national labour institutions and employment patterns. Most of these interventions take place with an array of unintended effects and, although some studies give examples of positive effects, in most cases the focus is on the unintended effects that limit CS's effectiveness overall. As Getz and Shreck (2006, p. 497) note, delivering the impact that was promised, given the 'realities of the global agro-food system, local production practices and consumer behaviour' can be 'a very ambitious goal indeed'.

4.5.2 Distributional dynamics

This section synthesises findings on distributional dynamics of certification benefits among certified producers along the lines of wealth, hired labour and gender. It is based on 97 studies. The number of studies contributing to each sub-section is reported on brackets.

Wealth and resources (n=20)

Producers' wealth and resources influence not only their integration in certified markets (see Section 4.5.1.1) but also the extent to which they can benefit from them. The term 'smallholder', is widely used by CS to describe their target population (i.e. Fairtrade, 2013b; COSA and ISSER, 2013; Rainforest Alliance, 2013), Nevertheless, the definition of 'small' producer can be problematic in some cases. For instance, Fairtrade defines smallholders as producers who are dependent 'on family labour as opposed to non-family labour' (Fairtrade, 2013b, p. 13), therefore allowing large producers which make use of only seasonal labour to enter Fairtrade markets and benefit the most out of the premia (which are linked to volumes as will be discussed further on), as Carimentrand and Ballet (2010) note.

In the case of Bolivian quinoa producers, Carimentrand and Ballet (2010), find that the dominance of the interests of larger producers led smaller producers to give up on membership in POs with Fairtrade certification or to be less likely to continue, feeling that their interests were no longer defended. Such findings are in accordance with Cramer et al (2014b) who report that the largest 'smallholders', i.e. the top 10% in terms of volume of sales to the cooperatives represented around 50% of total output sold in various cases of Ethiopian and Ugandan Fairtrade POs. Staib (2012) also reports on tensions between different wealth groups within the same PO, again illustrating that POs memberships are far from homogeneous.

Nevertheless, synergies between large and small producers are also possible, as the inclusion of larger producers is reported to help certified POs to reach the volumes

required by efficiencies of scale and market entry, thereby allowing them to continue offering market access and extension support to the smallest and most marginalised producers (Smith, 2010; Trauger, 2014).

While the extent to which the symbiosis of larger and smaller producers is truly beneficial for the latter remains debatable, it is clear that premia are linked to volumes, which are directly related with the amount of land available (Francesconi and Ruben, 2014) and thus benefit producers with higher production levels and larger plots more than smaller producers (Smith, 2010; Staib, 2012, Sutton, 2014; Valkila and Nygren, 2008; Cramer et al, 2014a; Cramer et al, 2014b). Findings from four case studies on Fairtrade bananas, for instance, suggest that the economic returns for producers with low volumes were limited, while producers with higher volumes were able to earn enough to re-invest in production (Smith, 2010).

Although Fairtrade POs tend to lack official policies for preferential treatment of the smallest producers, or a more equitable redistribution of benefits (Carimentrand and Ballet, 2010), cases where redistribution is attempted are reported. The most common form of redistribution, since it is part of compliance with standards, is the use of the premium (at least partly) for social development projects (Valkila and Nygren, 2008). The question, of course, is whether the investments funded by the premium benefit one group more than another within a PO, as cases of elite capture are often reported (Sutton, 2014; Cramer et al, 2014b; Staib, 2012; Dolan, 2010). Another way of redistributing is to use the price premium to pay additional amounts to poorest producers as Smith (2010) reports in relation to Fairtrade banana POs, where some steps were taken to (a) pay an additional amount to smaller producers (in Ecuador), or (b) all producers receiving the same price regardless of differences in quality, thus compensating for differences in access to irrigation, distance from processing facilities and quality of roads (in the Dominican Republic). These are instances where a pro-active policy of positive discrimination could help offset the potentially unequalising effects of certification.

Quality standards may also act as a barrier to entry for more disadvantaged producers. In this respect, larger and wealthier producers also appear to be better prepared to respond to stricter quality standards, while smaller producers struggle to overcome the structural barriers of the market and are likely to fall into a ‘quality treadmill’ (Shreck, 2002, p. 109). Getz and Schreck (2006), for example, report that in the case of Fairtrade-organic bananas, producers able to comply with higher quality standards were able to reinvest funds to further improve their quality, while the poorest farmers who tended to have lower quality bananas were less able to sell their product as certified and therefore received less benefits.

Access to inputs facilitated by POs and the additional benefits from certification may also be shaped by unequal power within certified POs, making, for example, wealthier farmers with higher production volumes command more and better services as they can use ‘their bargaining power to demand services on their terms’, while smaller and more remote producers tend to be neglected, as Staib (2012, p. 262) reports for Fairtrade-organic coffee producers in Nicaragua. This can happen in relation to both private and government subsidies (Cramer et al, 2014b). A similar

imbalance occurs between organic-certified and conventional producers, since the former have the resources to bear the extra costs of organic production and are also reported to receive more inputs and technical assistance than their non-organic comparators (Dowdall, 2012; Smith, 2010).

Several studies suggest that there is a direct relationship between higher volumes produced (and higher social status) and the power of a producer to influence and manipulate decision-making in certified POs, with important implications for the interests of the smaller producers (Francesconi and Ruben, 2014; Cramer et al, 2014b; Sutton, 2014; Sen, 2009; Fairtrade, 2013). For example, Cramer et al (2014b, p. 119), report on Fairtrade tea POs from Uganda where PO boards were disproportionately made up of shareholders 'distinguished by their wealth and education', while Fraser et al (2014) describe how a local elite in Nicaragua developed ownership over a Fairtrade coffee PO and excluded politically opposed producers, a pattern confirmed in the same country by Wilson (2010), who found wealthier producers more likely to occupy leadership positions within certified POs. Not surprisingly, elite control of certified PO boards has a knock-on effect on access to training and on decision-making related to premium use. For instance, board members are reported to benefit more from training initiatives (Mendez 2002), while the largest producers are also reported to receive the 'lion's share' of inputs subsidies and manipulate premia invested according to their own interests (Cramer et al, 2014b:119). These examples illustrate how difficult it is to empower more disadvantaged producers through more democratic practices in contexts where POs are embedded in deep-rooted power relations, manifested in various forms of elite capture.

Whether certified POs are controlled by the local elites or not, it is clear that since demand for certified is limited and not able to absorb supply, it is almost inevitable that larger and better prepared producers, POs and plantations are more likely to concentrate benefits. As already mentioned, Fairtrade importers tend to prefer working with bigger and better functioning POs (Pongratz-Chander, 2007), while the quota system, used by consortia and factories due to over-supply, is also reported to give preferential access to the certified market to larger producers, due to their greater reliability and capacity to deliver orders on time (Pongratz-Chander, 2007; Setrini, 2011).

In sum, in a context of farmer heterogeneity, including differentiation among smallholder farmers, the benefits of CS are unlikely to be equally shared, unless CS interventions radically alter the balance of power and existing inequalities. Most of the evidence reviewed here suggests that interventions of CS operate *within* existing structures of inequality and rarely substantially alter them, despite aspirations to democratise POs and their decision-making processes as well as the governance of large-scale agribusiness.

Gender

Producers and POs (n=24)

Gender equality is one of the aims in some of the social sustainability CS. However, breaking the barriers to gender equality is a task that in most cases requires sustained action and changes on many fronts. To what extent do the distributional dynamics of CS contribute to breaking such barriers? One question is whether prevailing gender divisions of labour in contexts where CS operate exacerbate or reduce the burden of work for women, or not. As discussed in Section 4.5.1.3, certified production, particularly when combined with organic, is reported to increase the workload of both male and female producers, though in different ways (Jaffee, 2006; Nelson et al, 2002; Lyon et al, 2010; Sen, 2009; Stenn, 2015). This gendered increase appears to disproportionately affect women's already heavy workload, as the domestic work burden remains unaffected (Hanson et al, 2012; Bacon, 2005 Lyon et al, 2010; Moberg, 2008). Evidence from some studies suggests that this higher burden is not compensated for by a distribution of benefits tilted in favour of women (Lyon et al, 2010; Sen, 2014).

Furthermore, although women may contribute significantly to certified production, their work tends to be less visible and less valued. Several reasons are reported. First, in terms of the division of labour men tend to get more involved in off-farm certification-related activities (i.e. meetings, technical training, planning, loans), while women are left with less visible (and more manual) tasks, such as plant care and post-harvest processing (Stenn, 2015; Nelson et al, 2002). Pongratz-Chander (2007) reports that women involved in certified coffee production in Costa Rica are considered as helpers rather than workers, while profits are controlled and distributed by the male members of the household. Second, women participating in certified production may not be officially registered with the certified PO due to land tenure issues (Sen, 2009), as many times women work on family land to which they may not have formal entitlement (Ellery, 2010; Nelson and Martin, 2013; Sutton, 2014; CESU, 2012; TWIN, 2013). Their exclusion may also be due to the cost of membership fees, as families are likely to pay fees only for one adult, which tends to be a male member (TWIN, 2013). Although membership and some benefits are reported to extend to the spouse (or daughter) of the registered producer (Laroche et al, 2012), women may not receive all membership benefits, such as training and extra income, or may not be able to vote in assemblies (Stathers and Gathuthi, 2013).

Whether officially members of certified POs or not, women involved in certified production are often reported to be disadvantaged in terms of both the benefits they receive and in their influence over decision-making within the certified POs. For instance, certification-related training may in theory be open to all PO members, in practice, however, women are reported to be less likely to participate, possibly because training is not tailored to their needs and agenda (Bergeron, 2012; Riisgaard et al, 2009; Fairtrade, 2015). Furthermore, general assemblies, as well as managerial staff and boards of directors among certified POs, are reported to be overwhelmingly male dominated (Lyon et al, 2010; Pollack, 2006; Sen, 2009), and female participation in leadership positions is very limited (Ronchi, 2002; CESU,

2012; Smith, 2004; Stathers and Gathuthi, 2013). Women are also commonly reported to be absent from decision-making regarding the use of the Fairtrade premium, resulting in projects that are more directly beneficial to men than to women, a fact attributed to traditional gender patterns that heavily restrict effective female participation in the key committees responsible for such decisions (Stenn, 2015; Ellery, 2010; Fairtrade, 2015; Said-Allsopp and Tallontire, 2014; Stathers and Gathuthi, 2013), but also to POs lacking explicit gender equity programmes or strategies (Hanson et al, 2012). Weak female participation in decision-making processes of certified POs is often linked to a lack of specific knowledge and skills (Terstappen, 2010), as well as low self-confidence, making women reluctant to speak about their experiences or attend trainings and meetings (TWIN, 2013; Stenn, 2015).

There are of course instances in which benefits for women may be greater. For example, one way to improve women's access to premium use, are women-only POs, which emerged as a mechanism of facilitating participation. However, they can also serve product and price differentiation, by creating a 'female-produced' brand that can command a higher price in the certified market (Hanson et al, 2012; TWIN, 2013). For example, TWIN (2013) reports on 'Women's Coffee' where coffee by female producers is separated out and receives an additional gender premium. Overall, women-only groups appear to improve female participation in organisational decision making, tailor premium and other investments to women's needs, as well as make women's work in commodity production more visible (Ellery, 2010; Pollack, 2006; Sen, 2009; CESU, 2012; TWIN, 2013; Stenn, 2015). Nevertheless, Stathers and Gathuthi (2013), also highlight that women-only groups may be more restricted by a lack of knowledge, communication and resources than mixed POs, and may therefore ultimately be damaging women's interests.

Plantation workers (n=12)

Regarding certified plantations, very few women are reported to hold supervisory or management positions, a fact that indicates that CS are limited in their ability to offset established socio-cultural norms which are inevitably reflected in the workplace (Loconto and Simbua, 2010; Smith, 2010; van Rijn et al, 2016; Nelson and Martin, 2013; Sen, 2009). Unlike van Rijn et al (2016), who attribute this under-representation of women in management positions to the lack of qualified women and therefore disassociate it from gender inequalities, Loconto and Simbua (2010) relates the scarcity of women in higher-level positions to the pre-certification male-dominated workforce, which, due to limited turnover, was still in place.

Evidence suggests that gender pay discrimination persists despite certification. Women workers are reported to earn less than their male colleagues, particularly if employed on temporary contracts (Smith, 2010; Walsh 2004), and to be more likely to be hired as temporary instead of permanent workers (Melkaraaen, 2009; Smith et al, 2004). Moreover, they appear to be subject to a variety of gendered patterns in terms of job allocation, with different remuneration scales for 'male' and 'female' jobs (Smith et al, 2004; Stathers and Gathuthi, 2013). Finally, women workers may also be excluded from accessing loans and receiving bonus payments, as in the case of certified tea plantations in Kenya (Stathers and Gathuthi, 2013).

In terms of non-monetary benefits, the Workers' Committees or Joint Bodies demanded by Fairtrade are reported to represent an opportunity for female participation (Nelson and Martin, 2013), however, the effectiveness of Joint Body (JB) as a mechanism of empowerment for women workers is questioned by Smith (2010), Said-Allsopp and Tallontire (2014) and Sen (2009). In fact, Said-Allsopp and Tallontire (2014) find that 'Gender Committees', a body instituted outside of Fairtrade interventions, ,and not Joint Bodies are more effective in producing the desired empowerment effects, as the latter tend to be male-dominated and may in fact partly offset empowerment. Finally, certified plantations are also reported to fail to do better than non-certified plantations in terms of sexual harassment, which appears to be largely viewed as "almost normal" and not necessarily something to be reported (CESU, 2012:28; Cramer et al, 2014a), suggesting once more that certification has not been able to offset the social context.

Wage workers (n=13)

A key finding about distributional dynamics among wage workers is that there is strong segmentation between workers hired by smallholders participating in CS and workers employed on certified plantations. In most cases workers hired by small producers are invisible in CS, compared to plantation workers, who are explicit targets for the monitoring of labour standards.

First, in many cases employees of certified smallholder farmers are simply not directly targeted by labour standards. This category is therefore not properly considered by the certification system, and, where they are, non-compliance with certification standards appear to be widespread (Cramer et al, 2014a; Shreck, 200; Trauger, 2014; Heller, 2010). Shreck (2002), for instance, reports there are no official FLO standards for hired labour employed by smallholders in banana production, but that the standards that apply for plantations appear only as recommendations for smallholder employers. Heller (2010p. 147), in the case of UTZ certified coffee in Guatemala, refers to a code of conduct regarding working conditions in smallholder farms which, besides being unclear on whether it is binding for employers or not, is also reported to be 'written with the large producers in mind' and not adapted for smaller ones.

It is not surprising that smallholder farm wage labour is not subject to the same scrutiny as plantation labour (Trauger, 2014), while, as already mentioned in Section 4.5.1.4, inspections are reported to miss non-compliance regarding working conditions of hired labourers working for smallholders (Heller 2010). It is certainly harder to target workers employed by small farms compared to plantation employees. In rural areas in many developing countries, agricultural labour markets are complex and differentiated, with families both hiring and hiring out labour (Jaffee, 2006). Moreover the 'disorganised, seasonal' nature of this labour force (Nelson et al, 2002:106) may contribute to its invisibility vis-a-vis CS. Despite these challenges there are a few instances of benefits trickling down to workers hired by certified smallholders, such as higher daily rates, paid social security, annual health checks, financial support for medical and educational costs, better accommodation and residency permits for migrant workers as reported by Smith (2010) and Nelson and

Martin (2013) for banana farms workers and tea pluckers, respectively. However, these reports also show evidence than not all workers involved in certified smallholder production appear to benefit from certification.

One argument is that smallholders are too resource poor and subject to volatile market conditions to offer decent work standards to their seasonal and casual workers. Thus these smaller employers may be unable to pay minimum wages or to offer other benefits (Shreck, 2002; Valkila, 2009, Heller, 2010; Trauger, 2014; Valkila and Nygren, 2008). This is especially clear in the case of unskilled workers, such as coffee pickers, for whom Dragusanu and Nunn (2014) find no evidence that Fairtrade increases their income. Trauger (2014) notes that Haitian (migrant) workers hired by Fairtrade banana producers in the Dominican Republic may not even know that they work on a certified farm, and Valkila and Nygren (2008) comment on the limited awareness of smallholder employers regarding Fairtrade standards for hired labour. Another characteristic example concerns Ghana's extensive sharecropping system for cocoa production, where caretakers provide labour on a continuous basis for the (often absentee) farm owners. Nevertheless, ‘they are not eligible for membership of Fairtrade and hence unlikely to receive benefits from it’ (Nelson et al, 2013:33). Nelson et al (2002), reporting on Fairtrade brazil nut producers in Peru, conclude that those who benefited the most were those already holding concession titles, while no clear benefits were detected for porters and shellers, the latter mainly women, and efforts to create an association and thus make their work visible have found no support or assistance from certification bodies. Finally, in a number of studies no differences in working conditions between workers working for certified and non-certified coffee processing plants or mills are detected, while ther are reports of labour law violations by certified coffee mills (Valkila and Nygren, 2008; Walsh, 2004; Cramer et al, 2014a). Finally, Valkila and Nygren (2008) comment that while farm workers may at least benefit from certification-funded community projects, workers in processing plants in urban areas cannot benefit from such projects.

There may also be unequal share of benefits among different groups of workers in certified large-scale plantations since temporary workers, as well as migrants and racially discriminated workers are reported to receive less benefits than their co-workers, creating tension among the workforce (Smith, 2010; Trauger, 2014; Smith et al, 2004). For instance, migrant or temporary workers are reported to be ‘almost systematically’ excluded from Fairtrade funded housing grants, while migrant workers are also restricted in benefiting from medical expenses coverage of their families or community investments (Smith, 2010). Furthermore, restrictions on using the Fairtrade premium to legalise the residency status of migrant workers, as in the case of Haitian workers in the Dominican Republic, not only is reported to leave these workers without any benefits from Fairtrade, but also to seriously undermine their ability to work (Trauger, 2014).

Conclusion

Overall, the synthesis of distributional dynamics points to the fact that in spite of the social objectives of CS, there are certain barriers that prevent smaller producers with lack of resources from reaping greater benefits from participation in certified

markets, while other groups are less visible, like female and wage workers working in smallholder farms, or migrant workers employed by plantations, despite their contribution in certified production. An important barrier to equitably sharing benefits is the differentiation that exists among producer members of POs, and the resulting differences in volumes of production sold affect the distribution of benefits. The most significant distributional dynamic among wage workers is the segmentation between workers employed by small farmers from certified POs and those employed by plantations. The latter are visible and directly targeted by the CS, whereas the former are invisible in most CS and therefore monitoring of labour standards is seldom applied to them. The evidence synthesised also suggests that CS have limited capacities to reduce structures of inequality based on gender, whether in relation to the visibility of women as producers or to their presence in key decision-making bodies within POs and plantation committees.

4.5.3 Contextual barriers and facilitators

Contextual factors play a major role in the performance of certification programmes and can act as barriers or facilitators to their effectiveness. Indeed, almost all CS acknowledge the importance of contextual factors, which limit the influence that CS interventions have on final impacts. As the causal chain linking interventions and final impacts (e.g. household income or health status) grows longer, the relative role of contextual factors increases (Nelson and Pound, 2009). This section synthesises findings on four key types of contextual factors: (a) the specific characteristics of POs, producers, and large-scale plantations; (b) the context of markets and supply chains; (c) the institutional and political context; and (d) wider socio-economic and environmental aspects. By far the most substantial evidence in qualitative studies is found for the first group of contextual factors. The section draws on 122 studies. The number of studies contributing to each sub-section is reported in brackets.

Characteristics of POs, producers and plantations

POs context: management, relationships with producers and buyers (n=46)

POs are the immediate context of certification programmes for smallholder producers, who still constitute a majority of certification participants in many CS in LMICs. Given that POs act as the key filter between the CS and the producers themselves in many cases, their strengths and weaknesses directly affect the effectiveness of CS.

As stated in the synthetic ToC in Section 1.3, good performance of POs is a basic assumption for CS to deliver benefits. Yet, cases of mismanagement and corruption appear to be recurrent, affecting both producers´ participation in certified markets as well as the degree to which they can benefit from them. The most common problems appear to be mismanagement and misuse of premia, high indebtedness, corruption, lack of transparency and the concentration of decision-making in management positions (Abarca-Orozco, 2015; Beuchelt, 2009; Lyon, 2005; Fraser et al, 2014; Milford, 2004; Milford, 2014; Smith 2007; Dolan, 2010; Francesconi and Ruben, 2014; Donovan and Poole, 2014). Such issues are of course interrelated, as one can lead to or enable the other. Additionally, they tend to be associated with producers'

limited ability to understand certification procedures and effectively control the management and board of directors of their PO (Beuchelt, 2009; Setrini, 2011; Milford, 2004; Staib, 2012; Donovan and Poole, 2014; Shreck, 2002).

This gap in information, knowledge and transparency between management and members enables the misuse of funds (Beuchelt, 2009; Setrini, 2011) and investments in accordance with the interests of PO management, but not necessarily those of the membership (Milford, 2004; Francesconi and Ruben, 2014; Sutton, 2014). These situations can thus generate mistrust towards the PO on part of members, thereby weakening the collective power of these organisations (Staib, 2012). Of course CS strive to offer various forms of capacity building support, through technical assistance and training, to prevent these situations, but this is often reported to be insufficient to improve PO management and empower producers-members to effectively control management's decisions and actions, as the gap between managerial requirements and existing skills is sometimes too large to fill (Fairtrade, 2013; Shreck, 2002). The other key tool used by CS, namely auditing systems, appear limited in their ability to prevent organisational and financial mismanagement (Beuchelt, 2009), and certification bodies are reported to intervene in some cases of mismanagement (Fraser et al, 2014), but not in others (Smith, 2007). Mismanagement and 'everyday governance' problems in POs are sometimes hard to detect through quick annual inspections that have several objectives to meet. Spotting these problems and taking corrective action takes times and resources that some CS may not have.

Despite substantial evidence of problems that prevent a more efficient and fair management of POs, it is clear that when POs work well, this is clearly reflected in greater effectiveness of CS interventions. Various studies have identified the following factors as facilitators of successful PO management: knowledgeable, skilled and experienced (though non-professional) leadership (Setrini, 2011; Sutton, 2014; Fairtrade, 2013); motivated, engaged and long-term committed professional staff (Beuchelt, 2009; Sutton, 2014); pro-activeness in seeking and securing extra credit, technical assistance, and funds from within and outside the CS (Bacon, 2005; Coop, 2012). Small PO size and consistent buyers' support (Sutton, 2014; Mendez 2002), as well as political and religious independence, are also reported to contribute in healthier POs and successful certification initiatives (Moberg, 2008; Fraser et al, 2014; Milford, 2014). Moreover, leadership rotation is reported to be an important mechanism for preventing management from taking full control of the organisation (Abarca-Orozco, 2015). Producers' capacity to question and hold their leaders accountable is also important, while fear, authoritarianism and clientelist relations may undermine such capacity (Setrini, 2009; Vasquez-Leon, 2010).

The effectiveness of POs hinges much on the relationship between PO management and key decision-makers and the base of the membership. It is possible that the interests of PO decision-makers may be different to, and even conflicting with, those of producer members, or at least with segments of the membership (Staib, 2012; Sutton, 2014). An example that illustrates this tension is the recurrent dilemma between retaining cash for investments and distributing profits to farmers, as Coop

(2012) indicate. Another example is the widely reported practice of side-selling, which can cause the PO important losses of market share, harming its interests and, by consequence, those of its members (Nelson et al, 2013). [89] Francesconi and Ruben (2014) report that, paradoxically, side-selling may be an outcome of the certification's own success in a PO. This happens when membership expands as a result of incentives offered by the CS through the PO, and this dilutes the magnitude of incentives to existing and new members (i.e. the distributed price premium or dividends) making the PO become a quasi-firm affected by principal-agent problems. Then side-selling becomes harder to control thereby eroding cohesion within the PO and incentives for members to continue selling through the certified channels. However, when incentives to sell production through certified channels improve, the unintended outcome may be that certified producers buy (rather than produce) non-certified products and deliver them to the PO as certified, causing an inflation in the volume of 'certified' produce that may not have followed the required standards (Heller, 2010).

In this complex relationship there are factors that can hinder or enhance producers' loyalty and participation, and, as a result, affect PO performance and certification effectiveness. Overall, transparency at the management level, but also during transactions with members (Beuchelt, 2009; Nelson et al, 2002; Stenn, 2015), as well as high-quality services, particularly credit and agricultural extension services (Donovan and Poole, 2014; Nelson et al, 2013; Ryan, 2011; Setrini, 2011), are reported to play a decisive role in attracting and retaining PO members. In particular, credit schemes are often reported as the prime reason for joining and remaining loyal to a PO, whether certified or not (Ryan, 2011), although Donovan and Pole (2014b) also report that indebted producers who are still in need of cash may side-sell their production, in order to avoid repaying their debt to the PO and secure some income for their families.

Furthermore, small PO size, as mentioned, appears to improve accountability and therefore PO-producer relationships (Sutton, 2014; Mendez, 2002). On the other hand, a larger size allows better access to export markets, although it also creates tension between reaching bigger volumes and providing high-quality services to more members, while the entry of professional managers of higher social and economic status is reported to create alienation between PO and members (Setrini, 2011; Sutton, 2014). There may be incentives for CS to target larger POs, partly to ensure a stronger base in a chosen location, partly to expand the number of associated members worldwide, which can have a positive impact on the marketing of the CS. What studies show is that PO size matters and it does so in somewhat contradictory ways that are very context-specific.

The relationship between POs and producers is also influenced by the origins of the POs themselves. Some studies suggest that externally imposed POs are more vulnerable to corruption and have weaker links with their members, while POs

[89] Side selling refers to the practice of PO members not selling their products to the PO but to other buyers.

formed on producers' initiative and efforts appear to have higher degrees and better quality of members' participation (Mendez, 2002; Shreck, 2002; Getz and Shreck, 2006). Interestingly, in the cases in which producers strongly identify themselves with their PO, certification status is reported to play a secondary role (Naylor, 2014; Setrini, 2011; Smith, 2010; Loconto and Simbua 2010) and may even be perceived with resentment as being told what to do by outsiders (Moberg, 2008). In other words, a CS would need to assess the implications of working with POs that have very different types of origins and relationships with their producers. Often this is not obvious at first glance.

Finally, producer-specific characteristics, both observable and unobservable, can also determine, at least partially, PO performance and certification effectiveness. For instance, producers' propensity to collaborate and support other PO members, is observed as a key factor of strong and successful POs (Jaffee, 2006; Dowdall, 2012), as are high self-confidence and low risk-aversion (Rueda and Lambin, 2013), characteristics that could also be related to higher education and an increased capacity to understand how certification works in comparison to the conventional market (Getz and Schreck, 2006; Terstappen, 2010; Shreck, 2002; Lyon et al, 2010; Balineau, 2011; Heller, 2010).

The relationship between POs and buyers is also crucial for the effectiveness of CS.[90] Included studies report on cases where benefits resulting from CS depend on the discretion of the buyers and the extent to which they are willing to go the extra mile for their suppliers, particularly in cases where CS appear limited in enforcing certain standards, such as timely and sufficient pre-payment (Walsh, 2004; see also Section 4.5.1.2 - Pre-payment and credit). For instance, cases of fully engaged buyers providing services and support to their suppliers beyond the stipulated standards are reported by Dowdall (2012), Setrini (2011) and Smith (2007). On the other hand, cases of buyers failing to honour their contracts and paying below the certified price are also reported (Abarca-Orozco, 2015).

Long lasting relations, but also direct and frequent contact and communication (Dowdall, 2012, Naylor, 2014; Mendez, 2002; Setrini, 2011; Terstappen, 2010) are reported to be crucial for solid and healthy PO-buyer relations, particularly when conventional prices rise above certified ones and a re-negotiation of contracts is necessary. Moreover, effective communication between buyer and producer is reported to enable the buyer to transmit knowledge and services to increase productivity and improve quality, and therefore producers' profits (Bakker, 2014; TWIN, 2012). Strong relationships of technical support and guidance with weak and inexperienced producers' associations, however, is also reported to lead to paternalistic relationships and greater dependency of the producing communities on the buyer (Shreck, 2002).

International buyers are often reported to provide incentives (Shreck, 2002) or actively lead initiatives (Abarca-Orozco, 2015; Smith 2007) to form POs in order for

[90] By buyer we refer to any actor further up in the value chain: exporters, importers, roasters, etc.

these to be certified. Buyers who are creditors to the PO are also reported to play a major role in restructuring the PO and in its decision-making (Donovan and Poole, 2014b). Furthermore, buyers are reported to support their supplying POs in various ways, from productivity investments and technical support to assuming losses or covering certification costs (Abarca-Orozco, 2015; Shreck, 2002; Heller, 2010; Lyon, 2005; Smith, 2010; Bakker 2014; TWIN, 2012). Such support is considered to increase the market knowledge and confidence of POs, enabling them to also approach non-certified buyers (Ronchi, 2002), and to upgrade in the value chain (Setrini, 2002).

Finally, if the main link is between buyers and producers, the certification itself may be an instrument that is sometimes deemed unnecessary. Thus, some studies report that buyers skip certification if it is not deemed beneficial, and offer benefits tailored to producers through personalised non-certified channels (Dowdall, 2012; Ellery, 2010; Smith, 2007). For example, Dowdall (2012) reports how US roasters encouraged their suppliers to drop the Fairtrade certification due to its high costs, and settle instead for a direct trade agreement on more favourable terms, thereby reducing the additional transaction costs of a Fairtrade certification. What many of these studies point to is the relative subordination of CS to the effectiveness of the business model in place, and whether a strong and effective relationship between producers, POs and buyers exists.

Besides the characteristics of PO management, their relationship with producers and with buyers, there are additional contextual factors that can hinder or enhance producers' participation in certified POS and CS effectiveness. The most commonly highlighted factors are the geographical location of the PO (Walsh, 2004; Schoonhoven-Speijer, 2012; Milford, 2014); appropriate investments in infrastructure that can improve quality product and help the PO differentiate itself in the market (Abarca-Orozco, 2015; Laroche et al, 2012; Nelson and Martin, 2013; Smith, 2007; Chiputwa, 2015); and producers' networks and consortiums that can enhance knowledge and experience sharing, provide better access to information, finance and markets, and improve bargaining power (Mendez, 2002; Laroche et al, 2012; Pongratz-Chander, 2007; Ronchi, 2002; Trauger, 2014; Pongratz-Chander, 2007; Smith, 2010; Riisgaard et al, 2009).

Plantations: management and workers' committees (n=10)

Just like POs, the specific characteristics of each plantation can affect the degree to which workers benefit from certification. Plantation management is very idiosyncratic and depends on an array of factors, including the location of the plantation, the origins of managers, the length of their presence in each location and many others. CS establish systems designed primarily to improve labour standards and the effectiveness of these control systems depends on the contextual characteristics of the plantations and their management.

The commitment of the plantation management to improve social and environmental conditions of production is often reported to precede certification, suggesting that systematic self-selection into CS may apply for plantations as well (Moore, 2010;

Raynolds, 2012; Schelly 2011). Moore (2010:28), for instance, states that regardless of certification status, some of the tea plantations he studied in India would still operate with the ‘same level of regard for workers rights’ due to the commitment of their management. Raynolds (2012), reporting on Fairtrade flower plantations in Ecuador, echoes this point, but further highlights that standards can provide practical guidelines to structure the good intentions and transform them into concrete social and environmental practices.

A lack of willingness to collaborate on part of management and administration staff, on the other hand, are reported to hinder workers’ empowerment processes (Lyall, 2014; Schelly, 2011). Similarly, a lack of workers’ knowledge of their rights and obligations, as well as of the certification mechanisms and benefits, means that workers cannot ascertain whether their rights are being infringed upon, and cannot protect themselves from abuses by management even within the certification framework (Lyall, 2014; van Rijn et al, 2016). Of course, a well conducted auditing process and frequent consultations with workers themselves could contribute to raising awareness among workers, enabling them to make claims through the established committees.

Workers committees, or Joint Bodies (JB), are supposed to provide a space for identifying workers' priorities and communicating with management (Raynolds, 2012; Schelly, 2011). Nevertheless, they are commonly reported to have limited decision-making power and capacity to address more controversial issues, such as inadequate compensation or pressure on workers to be more productive (Schelly, 2011; Sen, 2009; Moore, 2010; Brown, 2012). Brown, (2012:2585), for instance, describes JB as ‘a relatively safe [for the management] locus of worker participation’, with ‘intentionally limited’ interactions with unions, while Moore (2010:18) states that JB ‘re-legitimise the existing status quo’ without providing a real opportunity for workers to define and shape a space of effective participation. Makita (2012) and Sen (2009) support this view, reporting that workers in Fairtrade tea plantations in India were unaware of the JB members and activities, as well as of any certification-related funds, and that JB meetings were organised only on the occasions of important visitors, such as inspectors, researchers or buyers. Moreover, JB members often appear not to be elected by workers, but handpicked by management from the higher categories of the labour hierarchy and to have a rather inactive involvement in the JB, while workers from the lowest categories or those actively involved in unions are reported to be systematically excluded (Makita, 2012; Sen, 2009; Moore, 2010). Such control and manipulation of the JB can result in the management making substantial decisions over the premium use (Makita, 2012; Moore, 2010) and lead to workers’ failing to see their usefulness and regarding them as a facade (Schelly, 2011:95) or a ‘big waste of time’ (Sen, 2009, p. 219).

Finally, there are questions about the effects of certification on collective bargaining and trade unions, as it is suggested that that workers may have fewer incentives to bargain collectively if some of their needs can be met not through wage increases, but rather through certification premia (Brown, 2012). Brown further highlights that while collective bargaining involves all the represented workers, certification can only

benefit a fraction of the plantations of one region, and since it is voluntary, workers have limited say on certification-related decision-making, such as which farms get certified. Overall, even though trade unions can be heavily politicised and corrupt (Sen, 2009; Schelly, 2011; Melkeraaen, 2009), it is not clear whether JB can offer a better alternative, particularly since the latter are commonly reported to be far more susceptible to management pressure and limited in their collective bargaining capacity than the former (Raynolds, 2014; Moore, 2010).

Markets

Market Access(n=21)

Access to export markets is reported to be the main, and often the only, incentive for plantations and POs to adopt certification (Waarts et al, 2014; Setrini, 2011; Jari, 2013; Cepeda et al, 2013). While plantations appear more able to access export markets without certification, for POs certification may be the only way (Jari, 2013; Gomez-Cardona, 2012). Similarly, small farmers face the highest barriers to entry to more remunerative markets and their integration in POs through certification opens a door to new markets. At the same time, some standards have become compulsory for accessing certain markets, (i.e. GlobalGAP for UK supermarkets), thereby losing their voluntary character (and eliminating premia) (Dannenberg et al, 2013; Moberg, 2008; Ouma, 2010), while in some cases the industry is reported to shape standards according to its interests, as in the case of Unilever tea plantations and Rainforest Alliance standards (Loconto and Simbua, 2010).

Despite this power imbalance, certification is reported to improve access to export markets (Stathers and Gathuthi, 2013; Aidenvironment, 2016), which can be significant, particularly for the small producers whose only alternative are the very low prices offered in the local market (Getz and Schreck, 2006; Laroche et al, 2012; Setrini, 2011; Trauger, 2014). In fact, the lower the ability of local markets to remunerate, the bigger the advantage provided by certification to access export markets appears to be (Schreck, 2002; Smith, 2010). Certification is also reported to have been particularly beneficial in cases where producers had recently lost preferential market access due to changes in international trade regulations, as in the case of bananas from the Windward Islands (Smith, 2010).

Nevertheless, as discussed in Section 1.3, such benefits come with costs, which implies that certification provides a market advantage only to those who can afford to bear these costs (Schelly, 2011). Furthermore, gaining access to the certified niche market might involve significant effort, as fast changing standards may prove challenging to meet without technical assistance (Smith, 2007) and a few early entrants are reported to be controlling the market for certified produce, such as coffee (Valkila and Nygren, 2008).

CS alone may not be a definite factor facilitating access to export markets. Indeed, certification and business models are both reported to influence market access. Three examples illustrate this. First, certification projects implemented by companies and not NGOs appear to provide better market access as they have the possibility to

use their own supply chain and market structures (Fayet and Vermeulen, 2014). On the other hand, Chiputwa (2015) finds that Fairtrade producers, who hold their own certificate, had greater marketing freedom compared to Utz and organic farmers whose certificate was held by the exporter. Lastly, organic certification is reported to be a better market opener and to have a more assured demand, which is the reason why several cocoa POs in Ecuador are reported to quit the Rainforest Alliance certification in favour of organic (Cepeda et al, 2013).

Market demand and volatility (n=32)

The supply of certified products is widely reported to exceed demand by a large margin in many product lines, meaning that producers, POs and plantations are not able to sell all of their products in the certified market. Consequently, a substantial part of it ends up at the conventional market, even though it has been produced as certified (Beuchelt, 2009; Dragusanu and Nunn, 2014; Jaffee, 2006; Makita, 2012; Mendez, 2002; Moore, 2010; Nelson et al, 2002; Quaedvlieg et al, 2014; Nelson and Martin, 2013; Pongratz-Chander, 2007; Raynolds, 2012; Ronchi, 2002; Rotter, 1999; Schreck, 2002; Smith, 2007; Staib, 2012; Valkila and Nygren, 2008; Wilson, 2010; Moberg, 2005; Moberg, 2008; Stathers and Gathuthi, 2013; Fayet and Vermeulen, 2014; Minten et al, 2015). This problem appears to be central particularly in Fairtrade certification, where very few POs are reported to be able to sell the totality of their production in the certified market (Beuchelt, 2009), while for most the proportion ranges from 50% (Staib, 2012; Ronchi 2002; Valkila and Nygren, 2008; Mendez, 2002) to 30% (Wilson, 2010) or even 10% (Heller, 2010). Extreme cases where Fairtrade sales play a totally marginal role are also present, as in the case of a flagship Fairtrade tea producer, which is reported to sell only 5% of its production as Fairtrade (Loconto and Simbua, 2010).

Perhaps one of the main barriers in relation to market dynamics is that limited demand translates into limited economic impact on producers, as the certified price and premium are received only for the percentage of the volumes sold to the certified markets (Dragusanu and Nunn, 2014; Nelson et al, 2002; Pongratz-Chander, 2007; Luetchford, 2008b). Premia received for the percentage of volume sold as certified are often reported to be distributed equally or using a quota system across all the primary POs and all their members (Heller, 2010; Pongratz-Chander, 2007). Either way, as reported above, studies suggest that the premium gets so diluted that it becomes insignificant (i.e. Heller, 2010). Furthermore, limited demand is reported to act as a quality filter, allowing only producers who are able to deliver higher quality to sell greater volumes to the PO (see Section 4.5.2.1). Not surprisingly, therefore, guaranteed markets as part of a certification initiative are reported to be a key factor in successful certification, as Bakker (2014) and Herman (2010) highlight.

By contrast, in cases of increased global demand that exceeds global production, as is the case of cocoa, non-certified buyers are reported to offer similar services and benefits to producers as the ones offered by certification in order to incentivise production, causing certification to lose its comparative advantage over conventional markets (Ryan, 2011).

Moreover, market volatility and price fluctuation in conventional markets can greatly influence producers' benefits from certified markets. The case of certified coffee, a highly volatile commodity, illustrates how the comparative price advantage of Fairtrade-organic certification over the conventional market becomes larger when conventional prices reach their lowest level (Dowdall, 2012; Milford, 2014; Valkila, 2009). High conventional prices have the exact opposite effect, and as Valkila and Nygren (2008) report, during coffee price spikes the mainstream market can even offer better prices that the Fairtrade-organic POs. In such cases certified POs not only lose their comparative advantage over the mainstream market, but also members and market share, and risk defaulting on their contract obligations due to the reduced ability to purchase the required volumes (Beuchelt, 2009; Dowdall, 2012; Jaffee, 2006; Lyon, 2005; Valkila and Nygren, 2008). Members' loyalty in these cases is reported to be key for the survival of the POs (Lyon, 2005).

Overall, it is suggested, that, while during price crisis the Fairtrade floor price is really appreciated by producers (Valkila and Nygren, 2008), the failure of Fairtrade prices to adjust to rising international prices can turn the minimum price from 'lifebuoy' into 'shackles' for certified producers, and what was seen as a safety net can turn into constraint (Dowdall, 2012, p. 4). High fluctuations in mainstream prices can be a barrier to long-term buyers' commitment as well, and especially so for NGO-driven certification projects (Fayet and Vermeulen, 2014).

Local markets (n=8)

The institutional context of markets and local trade dynamics appear to be key determinants shaping certification outcomes, particularly in terms of price (Loconto and Simbua, 2010; Dolan, 2010). For instance, Dolan (2010) argues that the predominant structures of the tea auction in Kenya inhibit the partnerships that Fairtrade seeks to forge as the niche gets diluted with conventional output sold through the auction. Loconto and Simbua (2010:455), reporting on the Tanzanian tea industry, highlight that 'forms of relational governance that coordinate interaction among value chain actors', are more the result of pre-existing trading relationships, than driven by Fairtrade.

At the producer level, intermediaries (or middlemen, sometimes referred to as *coyotes*) also appear to continue playing an important role, even for certified farmers. Although they tend to pay less than certified POs (Bacon, 2005; Ronchi, 2002), they are valued for paying directly upon delivery, offering liquidity (advance payment, credit) during harvest, saving producers the transportation costs by purchasing at the farm gate, and accepting lower quality product that certified POs would reject (Bacon, 2005; Donovan and Poole, 2014b; Ronchi, 2002; Valkila and Nygren, 2008, Abarca-Orozco, 2015). As a result, even certified farmers tend to sell part of their produce to intermediaries as part of a strategy of market diversification and profit maximisation (Valkila and Nygren, 2008), a fact which contrasts with the Fairtrade 'publicity's rhetoric about cutting out the exploitative middleman', Walsh (2004:57) comments.

On the basis of the available qualitative evidence, as with power relations at local level and structural relations of inequality, the power of CS to radically alter locally-based market relations remains limited. Sometimes changes can happen at the margin but, overall, local trading dynamics tend to adapt to the arrival of CS while maintaining the core of pre-existing relations.

Institutional and political context (n=15)

Country-specific legislations related to POs and labour markets can facilitate or hinder the performance of POs and plantations, directly affecting their ability to access certified markets and return benefits to producers and workers. Local institutions and politics can hinder PO formation and performance and hence the impact of certification impact, especially when governance failures in national institutions are pervasive (Rotter, 1999; Dowdall, 2012; Sutton, 2014). On the other hand, the developmental trajectory of a PO may well be given impetus by effective government support, as in the case of Mexico (Abarca-Orozco, 2015) or Costa Rica (Pongratz-Chander 2007). In fact, democratic structures that allow producers to exercise pressure on their governments, coupled with strong state regulation, as in the case of the Ghanaian cocoa market, are reported to diminish the capacity of certification to make any real difference for producers, as Ryan, (2011:118) reports, concluding that 'Fairtrade did not bring about change in Ghanaian cocoa villages; Fairtrade has piggybacked on the democratic gains made by Ghana'.

Nevertheless, processes of market liberalisation, with the associated disintegration of regulatory agencies, as well as a lack of national quality standards appear to leave a void that certification initiatives can fill, potentially benefiting POs under pressure to offer more services to their members (Francesconi and Ruben, 2014; Naylor, 2014; Quaedvlieg et al, 2014). Similarly, the breakdown and absence of financial institutions or social programs aimed at supporting POs and producers add value to certification-related financial services, such as access to credit and pre-finance (Babin, 2012; Wilson, 2010), although as often reported such schemes may not be performing adequately enough to cover producers' needs.

Local historically-determined political power imbalances can also affect certification implementation dynamics and the benefits reaped. In some cases, certification is reported to apply a 'one-size-fits-all' approach instead of tailoring standards and implementation to the local context (Moore, 2012; Sen, 2009; Silva-Castaneda, 2012). For instance, Moore (2010) reports on the futility of requiring Indian managers and Gorkha workers to form 'democratic' JB, totally ignoring the power imbalance between the two groups, while similar incidents are reported by Sen (2009) regarding both gender and class power imbalances.

Overall, the evidence on the influence of national-level institutional dynamics and politics is not as substantial as one would have expected, given that the macro-institutional contexts do set the conditions for the formation, survival or decline of POs, as well as for the promotion of the kind of commercial agriculture that is then targeted by CS.

Socio-economic and environmental context (n=6)

In terms of socio-economic context, some historical zones of certified production are characterised by high labour migration, such as coffee producing areas in Mexico and Guatemala (Lyon et al, 2010) and the implications for certification may be significant. For instance, remittances from migrant members of the community can inflate the local labour market, increasing labour costs, and thus disproportionally affecting the costs of certified production, which is more labour intensive (Jaffee, 2006). Furthermore, Jaffee (2006) provides evidence that Fairtrade households send more migrants and receive more remittances, and therefore associates migration with sustaining the higher costs of certified production. Migration of male household members is reported to provide space for women to be more active in certified production (Lyon et al, 2010; see also Section, 4.5.2.2).

Moreover, certified producers often live in contexts of increased social insecurity and violence, and certification-related benefits may be even more valued for this reason. For instance, the provision of stable and formalised employment by Fairtrade certified banana plantations in Colombia was particularly valued by workers who lack legal employment alternatives in a context of drug trafficking and violence (Brown, 2012). Nevertheless, such situations are also reported to hinder producers' ability to participate in certified markets, such as female farmers in Mexico feeling too insecure to attend late meetings, and even facing death threats for their organisational activity (Pollack, 2006).

Finally, high national inflation rates can undermine the real relationship between the Fairtrade minimum price and FOB price and therefore should be considered when assessing Fairtrade benefits (Valkila and Nygren, 2008), while natural disasters, such as hurricanes, can radically affect producers' production capacity and in these cases producers are reported unable to reap the benefits of higher prices provided by certification (Dowdall, 2012).

Conclusion

This section on contextual factors has highlighted a wide range of aspects that act as barriers and facilitators, some very specific, others much broader. What they demonstrate is that CS do not operate in political, institutional, and social vacuum. Rather, they are incorporated into contexts where pre-existing relations and structures, whether related to the origin, trajectory or management of POs and plantations, or to the various markets from local to global level, or to the local and national social, cultural and political context, matter a lot and manifest forms of path dependency that make the effectiveness of CS highly context-specific. Showing that these contextual factors matter a lot does not imply that CS are irrelevant. In some cases, as shown in previous sections, they do reproduce pre-existing structures and relations, while in other cases they contribute to changing the rules of the game and therefore can shape context too. Overall, an implication for research and impact evaluations is that it is very hard to disentangle certification effects from contextual factors, and that a range of methods is necessary to reveal how the outcomes impact evaluations may observe are contingent upon configurations of context.

Table 17: Summary of findings of the qualitative synthesis (RQ2)

	Summary of findings - Implementation Dynamics
Targeting and (self-) selection of participants	Wealth • Capacity of POs, producers and plantations to bear the extra costs of certified production, withstand payment delays and obtain external financial support are key in CS adoption. • Production capacity, land size, household size, education and literacy skills, and degree of market integration influence participation in CS. • Findings point to systematic pre-existing differences in wealth and resources between certified producers, POs and plantations and non-certified or newcomers. Gender • Female participation in CS is undermined by difficulties in combining household work with certification-related activities, lower education and lack of skills, as well as socio-cultural norms. • Women-only POs may provide a more protected environment that enhances female participation.
Certification interventions and their implementation	Training, new farming practices and PO support • Tailoring training to producers' characteristics and needs, and providing sufficient and skilled extension staff are key elements of successful training. • A general lack of producers' knowledge and understanding of CS is reported. • Cost of applying new practices appears to be main barrier to adoption of GAP. • Certified POs tend to receive financial and technical support from a wide variety of actors, not always related to CS. Pre-payment & credit • The pre-payment standard is not always properly enforced. • CS appear to lead to better access to credit for POs, produces and workers. • Pre-payment and credit are often reported insufficient to cover costs of certified production. • CS-related credit gains importance during crop crises when other financial entities cease to support POs and smallholders. Minimum Price • The minimum price mechanism effectiveness depends on price volatility. During price slumps it allows POs to maintain their market share and provides stability for long-term investments, but becomes irrelevant during price spikes. • Oversupply of certified products affects the protection from the minimum price mechanism, as protection is related to the portion of production sold to the certified market. • Even when above the conventional price, the minimum price is often reported insufficient to cover costs of certified production. Price premium • Deductions in the final payment to the producer (PO administrative and certification costs, debt cancellations, price boosters) and oversupply affect the significance of the premium.

	• The collective management of the "social" premium can be undermined by local conflicts, lack of common interests, elite capture, and control and manipulation by plantations' management.
	Labour standards
	• The effectiveness of CS labour standards is related to local legislation and the degree to which it is enforced. Strong existing legislation may cause CS standards to become obsolete, while weak legislation, subject to frequent violations, may allow CS to bring significant improvements. • Overtime restrictions appear to be conflicting with the interests of workers and plantations, particularly when minimum wage is below the living wage.
Costs of certified production	• Production under organic-social standards significantly increases the use of hired and family labour. Poorer producers may rely more on child labour. • CS-related paperwork, meetings and trainings, are also reported to significantly increase producers' workload. • Stricter quality criteria and transition from conventional to organic production can cause product losses and therefore increase the cost of production. • Inspection and certification costs are reported to be significant and POs may need external financial support to deal with those.
M&A	• Internal monitoring systems may present opportunities, particularly for women, to acquire new skills and become actively involved in their PO. • Auditing bodies are reported to lack transparency and accountability towards producers, to be limited in grasping producers' reality and in making appropriate recommendations.
Spill-over and unintended CS effects	• CS appear to have an upward influence on local crop prices. • Spill-over effects are reported on the adoption of GAP by non-certified producers. • Increased requirements in labour of organic-social standards may increase demand in the local agricultural labour market. • CS create demand also for more "skilled" employment, such as auditors, extension staff, trainers etc. • Certification-related documentation is reported to be used to settle land disputes or to create "de facto" property.
Multi Cert.	• Overlaps between different standards can influence their effectiveness, both positively and negatively. • The dual Fairtrade-organic appears the most widespread combination. While organic certification is often required to access Fairtrade markets, particularly for coffee, increased costs of organic certification can be a barrier to Fairtrade adoption.
	Summary of findings - Distributional Dynamics
Wealth	• The term 'small' producer can be problematic, as it does not recognise that POs' membership base can be heterogeneous in terms of farm size. • Larger producers may benefit more from CS, as premia are linked to volumes, are more likely to comply with stricter quality criteria and may dominate the PO decision-making. But mutually-beneficial relationships between larger and smaller producers are also reported, as larger producers allow POs to reach the required efficiencies of scale.

Gender	• Female contributions in certified production tends to remain invisible, as female producers often lack the assets to register as PO members. • CS are related to a gendered increase in workload, which affects disproportionately female producers without an even distribution of benefits. • Weak female participation is commonly reported in decision-making related to PO management and premium use, as well as in supervisory/management positions in certified plantations. • Gender pay discrimination in certified plantations appears to persist despite certification presence. • The effectiveness of Joint Boards as mechanism of empowerment for women workers is questioned by several authors.
Wage workers	• Workers hired by small producers tend to remain invisible in CS and receive no, or very few, benefits. • In plantations, temporary workers, as well as migrant and racially discriminated workers, are reported to receive less benefits than their co-workers.
	Summary of findings - Contextual barriers and facilitators
Characteristics of POs, producers and plantations	POs context: management, relationship with producer and with buyers • POs' strengths and weaknesses directly affect the effectiveness of CS. • Cases of PO mismanagement and corruption appear to be recurrent, affecting producers´ participation in CS and the resulting benefits. • Producers' ability to understand CS and hold accountable their POs is key in CS effectiveness. • Transparency in management and transactions, good credit schemes and extension support are key in enhancing members' loyalty and participation. • PO size appears to matter, however, in very context-specific ways. Small PO size improves accountability, but is limited in accessing export markets. Large PO size allows better access to export markets, but losses in service quality and alienation between PO management and membership base are a risk. • Externally-imposed POs are more vulnerable to corruption and have weaker links with their members, while POs formed on producers' initiative and efforts have more and better quality members' participation. • Producers' propensity to collaborate, high self-confidence and low risk-aversion, and higher education influence CS effectiveness. • Long-lasting relations, direct and frequent contact and communication between PO and buyer contribute to CS effectiveness. • Engaged buyers may skip CS and offer directly benefits tailored to producers through personalised non-certified channels, if CS are not deemed effective. Plantations: management and workers' committees • Plantation management commitment to good social and environmental practices, as well as workers' knowledge of their rights and obligations and of CS mechanisms can enhance CS effectiveness. • Workers committees, or Joint Bodies, are reported to have limited decision making power and capacity to act, and may be vulnerable to management manipulation.

Markets	• CS are reported to improve access to export markets • Oversupply of certified products is a common challenge, particularly for Fairtrade, as an important part of certified crops end up at the conventional market. Guaranteed markets as part of a certification initiative are suggested to avoid oversupply. • Intermediaries continue to play an important role for certified farmers, as they pay directly upon delivery, can offer advanced payments, purchase at the farm gate and have lower quality criteria.
Context	• Local institutions and politics may hinder or enhance PO formation and performance and hence CS effectiveness. • Market liberalisation, disintegration of regulating agencies, and lack of national quality standards leave a void that CS can fill to the benefit of producers. On the contrary, democratic structures, ability to hold governments accountable and strong state regulation policies may limit the role that CS can play. • Local power imbalances can affect CS effectiveness and point to tailoring standards to the local context instead of applying a 'one-size-fits-all rule'. • CS benefits appear to be more valued in contexts of increased social insecurity and violence.

4.6 Integrated synthesis

4.6.1 Introduction

This integrated synthesis draws on the key findings from both the meta-analysis of included quantitative impact evaluations and the qualitative research with evidence on barriers, facilitators and broader contextual factors. The quantitative meta-analysis sought to address the question: What are the effects of certification schemes for sustainable agricultural production, and their associated interventions, in terms of intermediate and endpoint socio-economic outcomes for household/individual wellbeing, in low sand middle income countries? The findings about the barriers, facilitators and context contributing to explain differences in effectiveness are based on the synthesis of qualitative evidence used to address the questions: Under what circumstances and why do certification schemes for agricultural commodities have the *intended* and/or *unintended* effects? What are the barriers and facilitators to such certification's *intended* and/or *unintended* effects?

An important challenge for the integrated synthesis is the scarcity of 'linked' studies, i.e. of studies that contribute to both RQ1 and RQ2. Therefore, it is necessary to consider both sets of findings and extract key lessons from RQ2 studies that may help us understand effectiveness with sensitivity to context and heterogeneity. As we argued in Section 1.3, the causal pathways to impact for certification schemes are complex. We therefore draw on all the data available to us to illustrate where we have evidence of impact and how different contextual and other factors may act as barriers and facilitators to impact. From qualitative studies we learn a number of broad lessons mainly on implementation dynamics, distributional dynamics and external contextual factors. Whether they can be directly applied to the specific contexts of RQ1 studies cannot be clearly established in the absence of more substantive and specific evidence on implementation and context from these studies.

In this section we move along the causal chain, presenting main findings first on intermediate and then on final outcomes (see Section 3.1.4 for different types of outcomes considered in this review), considering the most relevant quantitative and qualitative evidence. The logic of the structure of this synthesis is the following. Given that many of the barriers, enablers and contextual factors affect multiple outcomes at the same time, we group the key outcomes in related causal chains (e.g. prices, yields, certified farm income) and address the contextual aspects that affect such group of causal chains.

Therefore, the section on intermediate outcomes (4.6.2) is organised around two empirically distinct sets:

- First, outcomes related to producers' farm profits and revenues, including key components, such as producer prices and yields, both key intermediate outcomes for many CS (Section 4.6.2.1). The discussion of quantitative effects is followed by a selection of key findings from RQ2 studies on implementation and contextual factors that could affect these outcomes.
- Second, labour market outcomes, mainly wages, for which some quantitative effects could be estimated (Section 4.6.2.2), followed by key findings from RQ2 studies on implementation and contextual factors that could affect these outcomes.

We then move onto consider evidence on final outcomes, mainly on household income, for which there is a respectable number of quantitative effect size estimates (Section4.6.3.1), and suggest key barriers and facilitators that mediate the effects on this outcome. This is followed by a selection of the main conclusions with regard to effects on other final outcomes, such as health and education, for which evidence was very limited indeed. Therefore, there is not much we can synthesise for that purpose, although some potentially useful points can be made on the basis of a rich body of qualitative literature on the causal chain linking some CS interventions to these other endpoint outcomes.

It must be stressed that the bulk on the quantitative and qualitative evidence synthesised in this report refers to *intermediate* outcomes and the causal chain between CS interventions and intermediate outcomes (yields, prices and farm income). There is much less that could be extracted to examine the full causal chain from these interventions through intermediate outcomes to final outcomes (total household income, assets, empowerment, health and education). Therefore, the main substance of the integrated synthesis is concentrated in the first section on intermediate outcomes. There is a general consensus in the literature that CS can be more directly linked to these outcomes and that the range of contextual factors is far too wide to establish a clear set of causal links, including barriers and facilitators, in relation to final outcomes.

By following the various hypothesized causal chains, we can extract a number of findings that can be summarised and illustrated in the form of a 'revised' ToC (**Figure 39** in Section 4.6.4), which attempts to provides insights into (a) what evidence we find on the key causal chains covered by the available studies; (b) where there is a

lack of evidence on hypothesised causal links derived from the original ToC developed in Section 1.3; and (c) which assumptions seem more critical to the different causal chains. In the final sub-section we draw our conclusions from the synthesis, before moving on to a wider discussion of our findings in Section 5.

4.6.2 Intermediate outcomes

Effects on producer farm yields, prices and farm income

In the ToC on the effects of CS there are three key intermediate outcomes for agricultural producers, all related. The most important is farm income, based on comparisons between treatment (certified) and control (not-certified) for any given commodity. Farm income depends on two main factors: the amount of output produced and sold and the price per unit. The producer price is quite close to being an input of the intervention, and can be considered an immediate outcome, especially in CS such as Fairtrade or Utz with direct price interventions via price premium.[91] However, in many cases the hypothesized additional price is the result of meeting standards that place the product in a more remunerative market. In the case of CS which do pay a price premium that may affect the final producer price (e.g. organic, Fairtrade, Utz, Rainforest Alliance among others)[92] the linkages between the premium and the final producer price are mediated by different factors, including the proportion of output sold to the certified market, the proportion of premium that is absorbed by PO or individual producer costs, compliance by buyers supposed to pay the price premium when mandatory and negotiations between buyers and producers over the amount of the premium. The price paid per unit of certified output may not matter as much as the amount of output produced for the certified market, and the output is usually a function of productivity. Therefore an important question is whether participating in a CS had any effect on yields. As discussed in Sections 1.2 and 1.3, not all CS design interventions to increase productivity but some certainly do, especially through capacity building actions as well as through incentives to invest in more efficient production commonly referred to as GAP – good agricultural practices (e.g. Utz, RA, MPS, GlobalGAP). Thus, the combination of effects on prices, yields and the amount of output sold in certified markets, which is determined by demand and supply constraints, leads to an effect on farm income. The following sub-sections summarise the main findings on quantitative effects on yields, prices and farm incomes, and the possible barriers and facilitators that account for observed effects and heterogeneity.

Yields

Five studies show a reduction in yields for certified farmers (SMD -0.42, 95%-CI from -1.23 to 0.39), although the effect is not statistically significant (p=0.312) and heterogeneity is very substantial (I^2=97.5%). For most individual studies the effect

[91] In the case of Utz thee price premium is mandatory but the amount negotiated between first buyer and producer. https://www.utz.org/wp-content/uploads/2015/12/Premium-Guidance-document-UTZ.pdf

[92] Consult standardsmap.org for comparisons between these CS on how price premium works.

size estimates are not statistically significant. Stratifying the studies by risk of bias produces no clear patterns. The CS concerned by the included studies (Fairtrade, Utz, RA, in isolation or combined with organic certification) are relatively similar in terms of the broad bundles of interventions. However, those that put explicit emphasis on productivity increases (Utz and RA) do marginally better than Fairtrade. In any case, effect size estimates for all three subgroups are not statistically significant. Overall, in terms of the causal chain in the revised ToC, this limited evidence suggests heterogeneous and inconclusive effects from capacity building interventions (considering the range of interventions favoured by the CS of these studies) and no evidence of positive effect from market price interventions (Fairtrade).

Final producer price

The meta-analysis findings for price effects indicated that there was limited evidence, but that this evidence points towards a positive effect of certification on prices. The overall effect is an increase in the price (SMD 0.28, 95%-CI from 0.08 to 0.47) and the effect is statistically significant (p=0.005). However, while there is less heterogeneity than for yields, it is still substantial (I^2=76.5%). Moreover, the forest plot suggests a pattern: there is a difference between schemes with a larger positive effect, where the additional price is associated with a quality premium or more demanding and remunerative markets (as for GlobalGAP), and schemes (Fairtrade + organic) in which the impact of the price premium on the final producer price is mixed. Overall, having only three quantitative studies showing some positive effect on prices is not sufficient evidence to support the price mechanism as an important contribution to farm incomes. Studies with high risk of bias produce a positive and significant pooled effect, while those rated moderate do not. The evidence of small but heterogeneous effects on prices as shown in the revised ToC remains in any case too limited to reach meaningful conclusions about this causal chain.

Farm income

Farm income is the outcome with the largest number of eligible studies. For clarity, what is being measured here is the effects of comparing the income producers receive from the production and sale of a particular certified commodity, such as coffee, with the income received by otherwise equivalent groups producing the same commodity but lacking certification. Therefore, it is only a partial picture of overall farm income effects. Whether the income from the certified production resulted in greater overall farm income was not established by the studies included in the meta-analysis as the comparison was made at commodity level.

The results from the 10 studies are mixed. The overall effect is positive and statistically significant, but not large (SMD 0.22, 95%-CI from 0.03 to 0.41). The degree of heterogeneity across studies is also very important as for the other intermediate outcomes. In fact, five studies show a positive effect while five studies had no or negative effect. In terms of differences between CS and their associated interventions, the only discernible pattern is that horticultural products certified by GlobalGAP generate larger positive effects than the other CS. For Utz there are

contrasting results from two different studies and effects in Fairtrade case studies are small and not statistically significant.

The heterogeneity in results for the three intermediate outcomes may be a function of the different methods used, as well as sample and supply chain characteristics. Some patterns appear in terms of differences between CS but these are not clear enough to warrant any strong conclusions. The period of time also matters as prices fluctuate and market opportunities may also be volatile, an important issue that looms large in some of the qualitative research discussed in section below. Generally, the observed heterogeneity in quantitative effects calls for a consideration of the contextual factors and barriers and facilitators revealed by the qualitative synthesis. Given the range of intermediate outcomes, their inter-relations, and the number of assumptions that need to hold for different bundles of interventions to have an effect on these outcomes, the next section extracts from the qualitative synthesis the key barriers, enablers and contextual factors that may explain some of the results found in the quantitative meta-analysis, and, more broadly, the mixed evidence on the impact of CS on these outcomes. This are the aspects for which this review found a more substantial body evidence, hence a discussion more detailed than for other causal chains is warranted below.

Accounting for mixed and heterogeneous effects in farm yields and incomes: barriers and facilitators in implementation, distributional dynamics and other contextual factors

As noted above, the determinants of farm income for certified producers compared to non-certified producers are many and these are combined in a variety of possible configurations, depending on the context and characteristics of the supply chain and markets in which these CS operate. The limited quantitative evidence synthesized does not leave room for generalisations but there is clear sense that heterogeneity is substantial and context matters. This section tackles the question of context in this group of causal chains highlighting the most important barriers and facilitators, and whether key assumptions in the ToC hold or not. There is not much evidence that directly links the quantitative studies with qualitative data, in the form of linked studies examining the same context, but these will be flagged below when relevant and available. Some of the barriers and facilitators selected below primarily refer to issues of implementation, i.e. how the actual practice of certification works vs how it is supposed to work. There are also issues of context and the structural characteristics of participants in CS and their inter-relations. The issues are organised following the key themes identified in the qualitative synthesis for consistency.

Costs of certification

Although the quantitative synthesis shows some studies with positive effects on prices and incomes (Weber, 2011; Minten et al, 2015, Subervie and Vagneron 2013; Becchetti et al, 2008; Riisgard et al, 2009), most qualitative studies tend to highlight the reasons why the additional price farmers obtain is not particularly effective at boosting incomes. Apart from the practice of deductions on premium payments,

reported by various studies (e.g. CESU, 2012; Milford, 2014; Valkila, 2009; Dowdall, 2012), which obviously reduce the extent of the price incentive, perhaps the most important issue is how certification generates additional costs. A large number of studies reported in the qualitative synthesis refer to direct certification costs, increased production costs, including additional labour requirements, which are induced by the process of certification, even by product losses related to stricter quality standards and restrictions on chemical use (e.g. Rueda and Lambin, 2013; Dowdall, 2012; Smith, 2010; Cepeda et al, 2013).

Capacity building interventions

Training for better farm management, better PO management and better agricultural practices is at the heart of many interventions in several CS. These interventions can impact on yields and certified farm incomes directly and indirectly. For example, Fairtrade focuses on training directed to empower producers to better engage with markets through their POs; Utz, MPS and GlobalGAP focus on interventions (technical assistance and guidance tools) that are likely to increase productivity as well as the quality of the final product. The training modalities may also vary, from extension services to training manuals to closer follow-up of practices at the individual level. The negative and statistically not significant quantitative effect reported on yields could suggest that such interventions are not sufficiently effective or that they are offset by other contextual factors. The reasons may also lie in lack of effective adoption of new farming practices because of increased cost, or unavailability of inputs. However, the number of quantitative studies is too limited to make such claims with any certainty. Moreover, most studies report on Fairtrade and organic certifications, which do not focus on yields.

Qualitative research on capacity building interventions, especially training, has raised questions over a number of key issues, as reported in the qualitative synthesis. Tailoring training to producers' characteristics and needs, and particular constraints (limited literacy, gender inequality) and providing sufficient and skilled extension staff are key elements of successful training, seem to be associated with immediate outcomes (knowledge, adoption of better practices) that could contribute to better yields, quality and incomes (Ellery, 2010; Pollack, 2006; Stathers and Gathuthi, 2013). Unfortunately, this is not achieved in many cases especially when interventions occur at level of POs with thousands of members. In this regard, the reported general lack of producers' knowledge and understanding of CS is cause for concern. Some examples suggest that approaching certification-induced training as 'development work' beyond a narrow focus on inspection and short-term technical support is seen as efficient in strengthening the institutional and financial capacity of certified POs, which in turn can positively impact farmers' profitability (Setrini, 2011:309).

Costs also matter and new farming practices entail additional costs that may not be affordable to all producers as suggested in the previous section. It is true that there may also be a trade-off between long-term yield gains and short-term increases in costs and workload, as in the case of some recommended practices, such as renewing plants or harvesting in shorter intervals (e.g. Arce, 2009; Lyon, 2005; Cofre

et al, 2012). These trade-offs across different time horizons may not be adequately captured by available studies.

Overall, even when well implemented, good training is not enough to secure better farm incomes and profits, and needs to be accompanied by more remunerative markets so that producers can adopt practices conducive to more sustainable farm incomes (Aidenvironment, 2016). The next section focuses on the dynamics of market access and trade relations, and associated barriers and facilitators.

Market access and trading relations

Different CS interventions can directly and indirectly impact on producer prices and certified farm incomes. Evidence on the effects on prices depends on the type of CS, as shown above. Different CS have different price policies. Some, such as Fairtrade, rely on a combination of a floor price guarantee and a premium to be used collectively, while others, like Utz, recommend price premia to reward achievement of standards and good agricultural prices, and other still operate by opening a door to more lucrative markets where demands are strict (examples include MPS and GlobalGAP). Whether they influence the final producer price or not depends on a myriad of factors. The qualitative synthesis has shed light on some of these factors. Below we highlight the ones that feature more prominently in affecting the causal chain between producer prices, market interventions and farm incomes/profits.

First, CS matter when they open up hitherto not accessible export markets. The vast majority of studies agree that, even after accounting for possible limitations and exceptions, certification improves access to export markets (e.g. Stathers and Gathuthi, 2013; Aidenvironment, 2016; Trauger, 2014). The question is whether (a) these export markets generate a rent via better trading conditions compared to local markets or alternative buyers (Schreck, 2002; Smith, 2010) and (b) whether producers targeted by CS are already engaged in export markets, and therefore benefit from a one-off structural change in their market options.

Second, the relationship between POs (producers) and buyers also matters, and whether CS impact on this relationship is a key issue *(*e.g. Jena, 2012; Subervie and Vagneron, 2013). Thus POs may get certified, but in many cases their relations with buyers predate certification and sometimes are not substantially affected by certification. In such a scenario the net effects on farmers' incomes and prices may be marginal. Another example is when engaged buyers offer directly benefits tailored to producers through personalised non-certified channels, if CS are not deemed effective, as in the case of "relationship" coffee (Valkila and Nygren, 2008).

Third, market demand conditions matter a lot. A general barrier that has been discussed in depth in the literature is the limited demand for certified goods, which translates into limited economic impact on producers, as the certified price and premium are received only for that percentage of the goods produced that are actually sold to the certified markets (e.g. Dragusanu and Nunn, 2014; Nelson et al, 2002). When this happens through POs, positive price effects from certification may dissipate with large numbers of members and limited certified proportion of output. A

large number of RQ2 studies, mostly reporting on Fairtrade, stress the extent to which certified POs and plantations only sell a fraction of their output, sometimes quite a low fraction, to certified channels (e.g. Beuchelt, 2009; Dragusanu and Nunn, 2014; Nelson and Martin, 2013; Ronchi, 2002; Valkila and Nygren, 2008; Minten et al 2015).

Finally, market price volatility and especially the occurrence of price slumps seem to be an important condition for greater impact of price interventions on prices and farm incomes. This is central to the effectiveness of the Fairtrade floor price/guarantee, for instance. In the case of certified coffee, a highly volatile commodity, the effectiveness of Fairtrade floor price interventions is corroborated when conventional price fall to low levels (Dowdall, 2012; Milford, 2014; Valkila, 2009). However, this intervention thus mostly acts as a mechanism of protection for bad times, which is of course important, without necessarily contributing to a sustained improvement in prices and therefore incomes.

The distributional dynamics of the premium and other certification rents

Aggregate average effects on farm incomes are partly affected by, and at the same time mask, important distributional dynamics that make the implementation of CS interventions not as uniform and equally shared as expected. This is an important issue for which quantitative evidence is lacking, but where qualitative evidence provides valuable insights. For those cases where the additional price is paid as a premium for a collective organisation such as a PO, some studies suggest that an important barrier for this premium to trickle down to individual producers is the combination of a limited proportion of their product going to the certified market (see above), and the distribution of the premium (net of any costs to be covered at PO level) to a large base of producer members. In the causal chain from capacity building, market interventions and funding for social investments, the translation of collective- level effects into outcomes on individual producers and workers also depends on the specific configurations of power relations affecting the management of POs. Power relations and local level inequalities affect both socioeconomic and gender distributional dynamics. The qualitative synthesis cites several studies that provide insights on premium use and distribution (e.g. CESU, 2012; Cepeda et al, 2013; Riisgaard et al, 2009; van Rijn et al, 2016), and a variety of cases of elite capture in the decision making of the 'social premium' are reported (e.g. Sutton, 2014; Staib, 2012; Dolan, 2008; 2010). The assumption that POs are homogenous groups of similarly-poor smallholders often does not hold. The reality is one where the structural inequalities commonly observed in rural areas also manifest themselves in the composition and power dynamics within POs. Thus, there is evidence that there is a direct relationship between higher individual production volumes (and higher social status) and the power of a producer to influence decision-making in certified POs, with important implications for the interests of more vulnerable producers (e.g. Francesconi and Ruben, 2014; Cramer et al, 2014b; Sutton, 2014). The implication is that large segments of smaller and poorer PO member producers reap a smaller proportion of the benefits accruing to the collective than their number suggest.

Studies reporting on gender dynamics also explore barriers to women obtaining a greater share of the benefits of certification in terms of farm income. Among the various explanations, two dominate the qualitative literature. First, although women may contribute significantly to certified production, their work tends to be less visible and less valued, because of the kinds of (manual) tasks they perform (Stenn, 2015; Nelson et al, 2002). In contexts characterised by patriarchal domination profits are controlled and distributed by the male members of the household (Pongratz-Chander, 2007). Second, inequalities in implementation processes are also reflected in studies which report on limits to female participation in POs. In particular, women are reported to be: (a) less likely to participate in training programmes that are largely designed by men for men (e.g. Riisgaard et al, 2009; Fairtrade, 2015), as well as in leadership positions within the POs (Ronchi 2002; CESU, 2012; Stathers and Gathuthi, 2013) and (b) often absent from decision-making regarding the use of the Fairtrade premium (e.g. Stenn, 2015; Fairtrade, 2015; Said-Allsopp and Tallontire, 2014). One possible facilitator of more benefits reaching women are women-only groups, which seem to improve female participation in organisational decision making, tailor premium and other investments to women's needs, as well as make women's work in commodity production more visible (e.g. Ellery, 2010; Sen, 2009; CESU, 2012; TWIN, 2013).

All these distributional challenges are not particular to CS interventions but a common reality of the contexts in which they operate.

Other contextual factors

As previous reviews of the literature have shown (e.g. Nelson and Pound, 2009), the qualitative synthesis in this report confirms that as the distance between interventions and final outcomes (e.g. household income or health status) widens, the relative role of contextual factors increases. In addition to the points made above on the dynamics of implementation of intervention and of distribution of benefits and costs, there are other contextual and external factors that can be classified in four groups.

- The characteristics and context of POs, producers, and large-scale plantations, which are rarely homogenous groups (as noted above).
- How markets and supply chains actually work (on this we have also already commented above).
- Existing policies and institutions that affect how markets and supply chains work, especially whether CS operate in an environment of deep market liberalisation, disintegration of regulating agencies, and lack of national quality standards or one characterised by democratic structures, ability to hold governments accountable and strong state regulation policies.
- Other contextual and more specific socio-economic and environmental aspects, affecting specific locations, supply chains and producers.

RQ2 studies provided evidence on each of these aspects but the first set of factors is the one that has attracted the attention of most researchers producing higher-quality qualitative work. Indeed, as discussed in the section on the use of premium, the

distributional dynamics and instances of elite capture are important features and remind us that the causal chain for many CS interventions implemented at the level of POs (and generally groups) includes first an effect on the PO/group and then the effect from the PO/group onto individual farmers. Therefore, the performance and accountability of POs is a basic assumption for CS to deliver benefits. The previous section has highlighted some of the features that particularly affect distributional dynamics.

For the purposes of illustration we focus here on the management issues and PO characteristics that matter most for effectiveness. First, a large number of studies analysed the governance of POs and instances of mismanagement and elite capture. These studies stressed gaps in information, knowledge and transparency between management and members (e.g. Abarca-Orozco, 2015; Milford, 2014; Dolan, 2010; Francesconi and Ruben, 2014). Sometimes the lack of alignment between PO leadership and members generates tensions, as with the recurrent dilemma between retaining cash for investments and distributing profits to farmers (Coop, 2012) or the widely reported practice of side-selling, which tends to increase as the membership of certification scheme expands, as in the case of Fairtrade reported by Francesconi & Ruben (2014). Of course, some interventions designed to strengthen POs' democratic practices as in Fairtrade, aim to address these issues, but the body of evidence suggests some of the problems are deep-rooted and hard to tackle through externally-driven capacity building or reporting requirements.

Second, a key issue seems to be whether POs and supporting organizations (NGOs, private business, buyers, etc.) provide credit and pre-payment, or not, and on what conditions. When pre-payments are on offer, POs are in a better position to buy the certified commodity before harvest, and therefore ensure the volumes needed to fulfil their contracts with buyers. At the same time producers receive cash when they most need it, enabling them to deal with the harvest costs, enhancing input use and in turn minimising side selling. RQ2 studies suggest that pre-payment is not as common as it should be (e.g. Milford, 2014; Walsh, 2004; Valkila and Nygren, 2008), and not always successful in delivering these benefits at either the PO or the producer level (Mendez, 2002; Riisgaard et al, 2009). There is some evidence that credit from certified POs to their members is provided on relatively better terms than credit supplied by conventional traders, especially compared to private money lenders and other middlemen (Milford, 2004; Nelson et al, 2002; Donovan and Poole, 2014; Dowdall, 2012). However, the effectiveness of credit provision also depends on the costs associated with certified production, which have been reported to be high compared to producers' gains from certification (Dowdall, 2012). Jaffee (2006) also shows that pre-payment and credit provided by certified POs is often used to cover the increased labour and input costs resulting from certification, thereby offsetting its potential beneficial impact. Credit from certified POs appears crucial especially in contexts in which there is a squeeze on conventional local credit providers, as for instance during the 2000-2004 coffee crisis in Nicaragua (Valkila and Nygren, 2008; Wilson, 2010).

Finally, well-performing key actors, especially within POs, are a central enabler of more effective interventions under CS. Thus the presence of knowledgeable, skilled and experienced leadership (e.g. Setrini, 2011; Fairtrade, 2013), and the availability of motivated, engaged and committed professional staff (Beuchelt, 2009; Sutton, 2014), enable PO actors to pro-actively seek and secure extra credit, technical assistance, and funds from both within and outside the CS (Bacon, 2005; Coop 2012). All of these factors contributed to cases in which the benefits in terms of a broad set of outcomes were noteworthy.

In sum, the causal chain between different types of interventions (capacity building, market interventions and funding for social investments), the final producer price, yields and certified farm income and profitability suggests there are several nodes where contextual factors matter, particularly in the form of barriers, according to the evidence we found for RQ2. Section 4.6.4 will recap some of these key factors along the causal chain in connection with the revised ToC illustrated in **Figure 39**.

Effects on wages and working conditions

The previous section focused on causal chains related to intermediate outcomes for agricultural producers. This section considers the evidence on labour standards and outcomes. The main intervention in this respect is the set of actions around monitoring and the verification of labour standards on certified farms, but intervention components linked to the collective (social or community) premium are also expected to impact on workers' objective wellbeing. Unfortunately, studies focused on labour outcomes are a minority in the review, and more generally in the literature, an issue already highlighted by previous reviews of the literature (Nelson and Pound, 2009). As noted in Section 4.1 the proportion of studies with a sole focus on wage workers is below 20% for both RQ1 and RQ2. Generally, the research interest in, and claims about, certification have centred on agricultural producers and, among them, especially on smallholder farmers. As a result, the evidence on effectiveness in terms of wages is very limited. There are eight studies with ES in the quantitative synthesis, four of which belong to the same research project (Cramer et al, 2014). These studies all report on wages and most also reported on other working conditions, but effect sizes for non-wage outcomes could not be calculated, with the exception of Ehlert et al (2014), who reported better access to training in certified fruit and vegetable production, and Schuster & Maertens (2014), who also report longer periods of employment for workers on certified farms. Results show some disparities between different sets of studies, especially a contrast between Cramer et al (2014, four studies) and Colen et al (2012), who report negative effects (statistically significant only for Cramer et al), and the other three studies, which find null effects. There are no studies with statistically significant positive effects. Overall, we find no quantitative evidence that the application of labour standards applied by CS contribute to higher wages compared to control groups. It is important to reiterate though that this synthesized result is based on a very limited amount of usable evidence.

Main contextual factors mediating effects of interventions on labour outcomes

If available quantitative effects are generally not positive, what factors may help us explain this result? There is a wealth of evidence from RQ2 studies without quantitative effects, but RQ1 studies on wages also contained relevant qualitative data on implementation issues, and particularly on distributional dynamics and other contextual issues. Here we focus only on the most salient factors. First, one problem is that not all workers are targeted. Despite the supposed importance of labour standards in many CS, there seems to be a lack of focus on different categories of agricultural wage workers, especially workers employed by small farmers (e.g. Cramer et al, 2014; Nelson et al, 2013; Waarts et al, 2016). Second, some studies suggest conventional labour standards may be too demanding as smallholders are too resource poor and subject to volatile market conditions to offer decent work standards to their seasonal and casual workers (e.g. Shreck, 2002; Valkila, 2009; Heller, 2010; Valkila and Nygren, 2008). However, Cramer et al (2014a), included for RQ1, report worse working conditions in certified small farms compared to non-certified small farms, so scale is not the main confounding factor. Third, monitoring and auditing working conditions in smallholder production areas is logistically impossible for many CS, who may be operating with POs that have thousands of members scattered across wide areas. Thus, for example, inspections are reported to miss non-compliance regarding working conditions of hired labourers working for smallholders (Heller, 2010).

Nevertheless, the lack of evidence for positive contributions of CS to wages cannot be simply linked to the invisibility of wage workers employed by smallholders. There are other factors, since a lack of evidence of positive effects on wages also concerns plantations. Much of the qualitative evidence points to the significance of the context of national labour institutions. Thus, in cases where national labour laws cover and exceed the guarantees offered by CS, and are properly enforced, benefits from standards become irrelevant (e.g. Moore, 2010; Valkila and Nygren, 2008; Nelson and Martin, 2013; Raynolds, 2012). Conversely, some studies show that in contexts where labour legislation and its enforcement are weak there is more scope for improvements in certain aspects of working conditions (e.g. Smith, 2010; Raynolds, 2012, 2014), even though we do not have quantitative evidence for these instances. With regard to non-wage conditions several studies also suggest that plantation management already applied comparably high labour standards prior to certification, suggesting that systematic self-selection into CS may apply for plantations as well (e.g. Raynolds, 2012; Schelly 2011).

Finally, especially for certified plantations, minimum wage requirements (one of the core labour standards) align with nationally stipulated minimum wages, which are often too low to cover living costs, or too low compared to ongoing average wages, so it is not surprising there is no premium in wages unless employers have a stronger incentive (Raynolds, 2014; Schelly, 2011; Smith, 2010). A response is manifested in the Global Living Wage Coalition, which includes six CS that are ISEAL members, which ‘recognise that a living wage is crucial to their individual certification programmes and they have agreed to a shared approach for measuring

living wage'.[93] Given the evidence on the difficulties of successfully monitoring labour standards, including minimum wage payments (Moore 2010; Schelly 2011; Heller 2010), efforts to agree on a living wage may be met by resistance from employers, more so if they disagree with the way the living wage is calculated.

Wages are of course not the only important outcome related to labour standards but it was the only one for which there was more than one study available for quantitative meta-analysis. Some of the quantitative, as well as the qualitative, studies also report on other working conditions such as benefits, health & safety issues and workers' empowerment. In this regard the evidence is mixed. Some studies report worse conditions across a number of labour standards (Cramer et al 2014; Luetchford 2008). However, one of the main advantages of CS is the greater visibility of certified plantations to auditors and pressure groups, and the greater pressures for better labour conditions sometimes present in certified-driven value chains (Smith, 2010; Raynolds, 2014).

Another contextual factor concerns the dynamics of collective bargaining and the role of unions and CS-established committees in plantations. These aspects can be very important for the achievement of improvements in working conditions in plantations. Whereas in many cases CS are not very active on this front and limit themselves to requiring basic standards such as freedom of association, some CS have been pro-active in steering organisational interventions for this purpose, as in the case of Fairtrade and its Workers' Committees, or Joint Bodies (JB), discussed in the qualitative synthesis. An important barrier seems to be the limited decision making power of these committees, and their weak capacity to address more controversial issues, especially payment and working conditions grievances (e.g. Schelly, 2011; Moore, 2010; Brown, 2012). Another issue is whether Joint Bodies can or should operate as alternatives to established trade unions, given that they are far more susceptible to management pressure and more limited in their collective bargaining capacity than the unions, especially in contexts of more conflictual labour relations (Raynolds, 2014; Moore, 2010).

Overall, the evidence we have found on quantitative effects on working conditions is limited, but does suggest that CS seem to face obstacles to producing positive impacts on wages, a key labour market outcome. We have highlighted several factors that help us understand the relative ineffectiveness observed for the limited number of studies available for this outcome. Among them we can stress: the role of national context for labour institutions and predominant labour relations, which sometimes make certification labour standards irrelevant; the fact that labour standards are only applied to a segment of the population of wage workers, generally excluding those employed by smallholder farmers; the incorporation of new collective bodies such as worker committees lacking bargaining capacity.

[93] http://www.isealalliance.org/our-work/improving-effectiveness/global-living-wage-coalition

4.6.3 Final outcomes

The causal chain to household income and assets

Turning the final outcomes, there were eight studies available for meta-analysis dealing with total household income and two only for assets (wealth). In the case of household income the synthesised effect is positive but not statistically significant (SMD 0.13, 95%-CI from -0.06 to 0.32). While there were large variations in effect sizes across studies, sensitivity analysis suggests the result is robust. As for other outcomes, heterogeneity is significant despite the fact that the majority of included studies deal with Fairtrade certification, including the two studies with the most contradictory findings.

In the case of assets and wealth, the evidence is even more limited, i.e. from just two studies. Hence, with so few studies the pooled effect has only limited value. The joint effect is a small increase in wealth among certified producers (SMD 0.05, 95%-CI from -0.15 to 0.26), but the effect is not statistically significant (p=0.598). For both income and wealth the results are not sensitive to the inclusion or exclusion of high risk of bias studies. In sum, the null effect for both outcomes is very robust, at least on the evidence available.

In the hypothesized causal chain effects on overall household income and assets will depend on the dynamics of effects on farm and labour incomes associated with certification, so interventions directly affecting farm incomes are likely to impact on household income and assets. There may also be linkages with community-level investments funded by CS (through premium or credit), which may impact on assets. Most of the barriers and facilitators discussed in Sections 4.6.2.1 and 4.6.2.2 are also relevant to household incomes and assets as final outcomes. In addition, it is necessary to consider the role of the following factors:

- The degree of reliance on certified farm income
- The linkages between certified income and other sources of income
- Off-farm employment opportunities
- Other external forms of support (family, organisations)

Most studies included for RQ2 provide information on the first of those factors, suggesting that in some cases producers depend on other sources of income and therefore a marginal positive effect on certified farm income is unlikely to substantially impact on total household income. The causal chain between certified farm income and labour income from certified production therefore depends on the relative dependence of households on these sources of income, and on what different household members do and how they contribute to household revenues. If self –selection of more prosperous producers or better-managed POs occur obviously household income effects may be biased if no adequate counterfactual is considered (e.g. Asfaw et al, 2009; Minten et al, 2015; Nelson et al, 2013;). To an extent, RQ1 studies deal with selection bias problems, but they can do so only within particular technical parameters which concern only the estimation of impact. The fact remains that richer and larger famers tend to be over-represented among certified farmers, something established by most studies in the review.

Other outcomes: Health, education and empowerment

The evidence on health and education outcomes for both the quantitative meta-analysis and the qualitative synthesis was very limited, to the point that not much that is meaningful can be said, with the exception of effect sizes for schooling. There were two studies that reported on measures of health (an illness index) and five on schooling. The pooled effect size for schooling is positive and significant but small (SMD 0.12, 95%-CI from 0.00 to 0.24), mainly due large differences between two sets of studies: three studies with a null effect and two with positive and statistically significant effects, especially Bennet et al (2012) who report on RA in cocoa in Cote d'Ivoire.[94] Interestingly, for schooling the studies rated as moderate risk of bias return a positive and statistically significant effect, while the high risk of bias ones, as a group, do not.

We found no studies that met the inclusion criteria for RQ1 and reported on measures of empowerment. However, studies included for RQ2 contained substantial qualitative evidence on possible barriers and facilitators to bringing about empowerment outcomes. These are summarised below.

Due to the multiplicity of contextual factors at play, the incidence of multi-certification, the proliferation of different institutions giving support to participants in CS and the self-selection of better-off producers into CS, it is very hard to establish any meaningful connection between certification and its associated interventions and improvements in health and education outcomes. According to the ToC it is possible to think of a number of hypothetical linkages:

- Positive effects on household income result in more investment in education and more resources for health through basic expenditure effects, assuming the additional income is spent and/or invested
- Good agricultural practices, especially when protecting producers and workers from harmful inputs and ways of working could have a direct impact on health
- The premium in Fairtrade schemes can lead to investments in infrastructure that improves access to health and education.

These linkages remain plausible, but it is remarkable that few studies have reported on such outcomes or tried to analyse the key nodes in these causal chains. For the qualitative synthesis we were unable to identify any substantive evidence on barriers and facilitators for health outcomes. Likewise, the main substance of evidence related to education and schooling in included qualitative studies referred to the role of education in self-selection, i.e. the extent to which certified farmers, especially in CS with more demanding standards (e.g. GlobalGAP), were able to enter these markets and obtain certification precisely because of their higher education levels (Cofre et al, 2012; Dowall, 2012).

[94] If unit of analysis adjustments are applied the pooled effect for schooling remains positive but is no longer statistically significant, see Annex F.

Evidence of elite capture in investments undertaken by POs, especially in the use of Fairtrade premium, discussed already in section above, may also impact on the distributional effects of investments in education and health facilities, if the poorest segments of the population in the areas under certification fail to benefit because of the existence of user fees, for example (Cramer et al, 2014). Therefore, it is not possible to establish what facilitators may have underpinned the positive results of two of the included RQ1 studies, besides indirect expenditure effects from household income and wealth, or the fact that better-off farmers, who are more likely to adopt standards like RA, may be more prone to invest in the education of their children.

With regards to empowerment, a substantial body of evidence for RQ2 suggests that the presence of external support is critical to initiating and maintaining both certification-related projects and POs, as well as possible opportunities for empowerment individually and collectively (e.g. Bakker, 2014, Herman, 2010; Stathers and Gathuthi, 2013; Arce, 2009; Lyon, 2005; Pongratz-Chander, 2007; Quaedvlieg et al, 2014; Smith, 2007; Nelson et al, 2013; Fairtrade, 2013; Roy and Thorat, 2008; Ouma, 2010; Fairtrade, 2013). This raises questions about the sustainability of certification in the absence of systematic external support, whether coming from NGOs, the CS themselves or other partners, including governments and businesses (buyers). Studies reviewed for RQ2 also raised doubts about the effectiveness of some CS interventions in terms of empowerment of women and wage workers, as discussed in other parts of this section (e.g. Lyall (2014; Schelly, 2011; Smith 2010; Said-Allsopp and Tallontire, 2014).

Empowerment may indeed come as a result of access to new and more remunerative export markets, as discussed above. However, the distributional dynamics of this access and the mediating role of POs, plantations and buyers may mean that not all producers or workers are equally empowered, and that sometimes producers may not be even aware that they have enhanced market power as a result of certification and thus fail to make use of it. The limited demand for certified products and the fact that many POs still sell significant proportions of output to conventional channels puts a question mark on this linkage, which is plausible but not guaranteed.

4.6.4 Towards a revised toc

What do the results reviewed above tell us about the synthetic ToC proposed in section 1.4 of this report? It is clear from the quantitative synthesis that there is a general absence of high-quality evidence on effects, i.e. a lack of a sufficient number of low-RoB studies for most intermediate and especially for final outcomes. The limited meta-analysis that was possible given the body of evidence did suggest significant heterogeneity and no clear-cut conclusions about which CS are more effective or what ‘bundles’ of interventions could generate more positive effects on different intermediate and endpoint outcomes.

Therefore, it is not possible to draw substantial insights into the ToC that informed this review, in terms of what specific kinds of interventions or CS are more effective. The original ToC drew on what most CS do, in terms of bundles of interventions, to achieve a series of intermediate outcomes, which the CS normally link to much

broader (and ambitious) sustainability goals.

There are no major differences between the hypothesized ToC and the one presented below as part of this integrated synthesis, at least in terms of the expected causal pathways. The main differences are the following. First, we consider more specific outcomes in the sequence which they follow according to the evidence collected. Second, the set of assumptions is updated to reflect what the body of knowledge reviewed highlights as key conditions for effectiveness. Third, the linkages between bundles of interventions and different sets of outcomes are less linear and reflect more lines of actual or potential causality. Fourth, the updated ToC reflects the lack of evidence for many of the linkages considered, both between interventions and intermediate outcomes and between different sets of outcomes (intermediate to final). The quantitative evaluations we have been able to synthesize only provide limited, often inconclusive and generally heterogeneous evidence on the main linkages between types of interventions and intermediate outcomes. The remainder of this section summarises key points on the main causal chains as illustrated in **Figure 39** below.

As argued in section 1 of this report, each CS comes with a bundle of interventions and it is hard to disentangle which interventions matter most on the basis of quantitative evaluations that use the fact of being certified as a proxy for receiving bundles of interventions. In other words, most included studies estimate effect sizes from 'certification X' (or even bundles of certifications) on selected outcomes.[95] Therefore, it is in many cases virtually impossible to tell whether an effect size suggests that a particular type of intervention is effective or not; we can only say whether being part of a CS has any impact on the outcome. From the CS included we can infer which bundles of interventions dominate. As shown in Section 1 of this report, some CS are more focused on market price-type interventions and PO support (such as Fairtrade), whereas other CS are more demanding in terms of labour standards (e.g. MPS-SQ), and some add capacity building interventions to improve productivity (Utz, RA). However, we have also observed significant overlaps between CS in terms of the number of standards and areas they cover (Section 4.1).

If we start from the causal chains affecting yields, we start by considering capacity building interventions. First, the evidence on effects on yields is mixed and in some cases the effects are negative, but many of the interventions (e.g. from Fairtrade) are not designed to expand yields. CS that use capacity building to improve yields have generally slightly but statistically insignificant effects on yields (Utz, RA). Second, CS which use other market price interventions or a different focus on capacity building (towards empowering POs, strengthening their position in the value chain), such as Fairtrade, are associated with lower yields, although yields are not a focus of their interventions. It is impossible to establish whether market-type interventions have any positive effect on yields, given the heterogeneity of results and types of interventions. Fairtrade certification, associated with price interventions, seems

[95] A number of studies do separate out the impact of certification from the impact of belonging to a PO through the use of appropriate sampling designs. Other studies use sampling design to isolate the impact of training, for instance.

associated with lower yields, but that does not mean that the price incentives do not work, as their effect may be offset by other aspects, such as, for example, adoption of organic standards. On the basis of qualitative studies, the key barriers and enablers affecting this chain are:

- For capacity building to yields:
    - Capacity building is adequate and tailored to context (not always the case).
    - Producers can afford the cost of adopting standards (which often they are not).
    - POs have sufficient and sustained external support to adopt and maintain standards, for which evidence is mixed.
- For market interventions to yields:
    - Pre-payment is enforced and sufficient (problems of implementation frequently reported)
    - Credit is provided, is sufficient to cover production costs, and is efficiently used (not many CS provide this and it is not always used for productivity improvements)

The evidence for the causal chain between market interventions and producer prices is also limited to a few studies, but effects seem generally positive. Whether this is because of price guarantees, price premium (Fairtrade) or tapping into more remunerative markets because requirements are met (GlobalGAP) is not possible to establish, although the quantitative effects seem larger in the latter case. The main assumptions for this chain, according to the RQ2 studies are the following:

- Sufficiently remunerative markets, i.e. conventional prices are below the floor price and/or the price premium is enough to compensate for additional costs (many studies showing this only happens in certain conditions). Indeed costs of certification have been stressed as a key barrier to adoption and effectiveness

Overall price interventions may have some support, especially when we consider floor prices in contexts of price slumps. The fact that certain CS open a door to more lucrative markets also has an effect on the final producer price.

The causal chain affecting farm income works as a combination of the previous two chains, but some comments are in order. First, we find that the overall synthetic effect is positive but highly heterogeneous. Judging by the CS that seem to do better (GlobalGAP and Utz, though very mixed in the latter), the result may be due to a combination of more effective capacity building for productivity increases, combined with more remunerative markets, partly because of the type of commodity (higher value-added horticulture vs coffee or cocoa), and partly because of the quality demands associated with these CS. The evidence from studies on Fairtrade is less clear and effect size estimates are not statistically significant, which might be the result of a combination of higher prices, lower or similar yields and a small market. Qualitative research reviewed for RQ2 suggests the following assumptions being critical for this chain, in addition to the ones mentioned for prices and yields.

- Markets are able to absorb certified products (demand constraints frequently mentioned as a barrier to growth in certified farm incomes, as discussed in previous sections).
- PO-buyer communication is effective and trading relations more beneficial to certified producers (CS not always changing pre-existing trading relations).

The revised ToC also revisits the basic link between labour standards interventions and labour outcomes. Our quantitative evidence does not supporting such link, but the number of studies concerned is extremely limited. This is one of the areas where rigorous quantitative research is sorely needed. According to RQ2 studies, the critical assumptions in this causal chain, between interventions, labour outcomes and effects on household income and social outcomes, are: (a) all workers are targeted (including those employed by smallholders); (b) labour standards exceed national laws and are properly enforced. The former assumption does not hold as labour standards are often only applied to a subset of participants in CS. The latter depends much on each national context, but available studies have reported cases in which the assumption does not hold. This is a highly contextual factor that is outside the control of most CS, unless they succeed in interventions towards enforcing a 'living wage' as some are trying to do.

On the causal chain between social investments (whether funded by the Fairtrade 'social' premium or other forms of support) and health and education outcomes, the findings are inconclusive. The only cases with statistically significantly positive effect size estimates concern RA and other CS and not Fairtrade. Nonetheless, various studies included for RQ2 provide examples of positive linkages, especially with schooling, but, as argued in the qualitative synthesis, it is not clear that the distributional effects are neutral. Therefore, given the lack of evidence, this remains a hypothesized relationship, and one that is plausible if distributional dynamics are seriously considered, i.e. if it is recognised that 'communities' are not as homogeneous as often assumed.

Overall this revised ToC confirms the expectation of the original ToC (Section 1.3) that we are dealing with multiple and partly overlapping theories of change, and therefore bundles of interventions from which it is almost impossible to disentangle the effects of specific interventions - a nexus array of possible linkages that can be hypothesised for which we have limited or no evidence at all. The synthesis has shown that we find complex bundles of interventions with some variations of emphasis between CS. To complicate matters further these may not be consistently implemented, depending on context. The qualitative synthesis has provided evidence on a substantial number of contextual factors, and issues of implementation and distribution, that are critical to understanding the complexity of causal chains considered for an effectiveness review of certification in agriculture.

4.6.5 Authors' conclusions

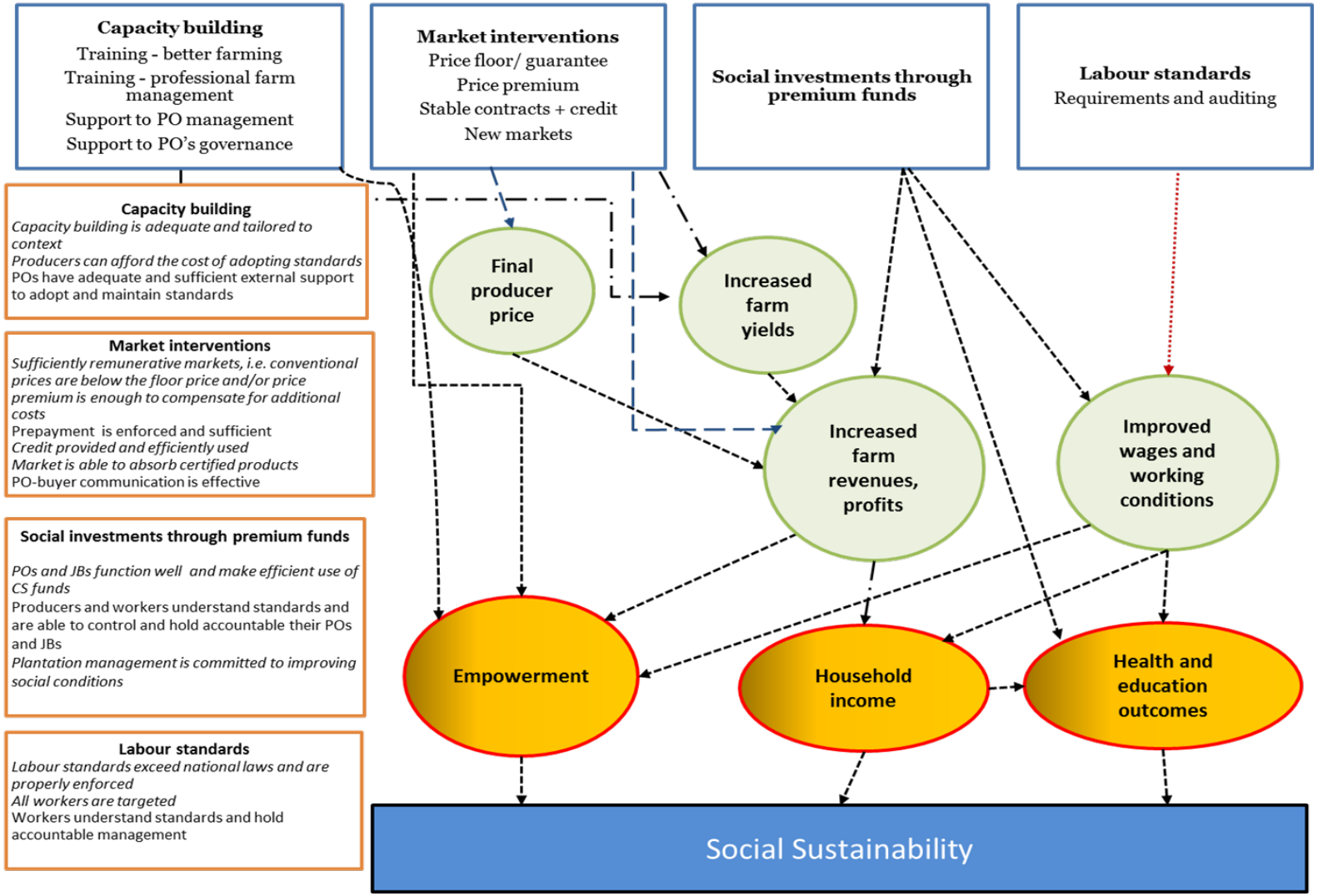
Overall, the general conclusion for this section and also for the review is that there is an absence of evidence of effects of certification of agricultural production on socio-economic outcomes of producers and workers. There are many outcomes, multiple

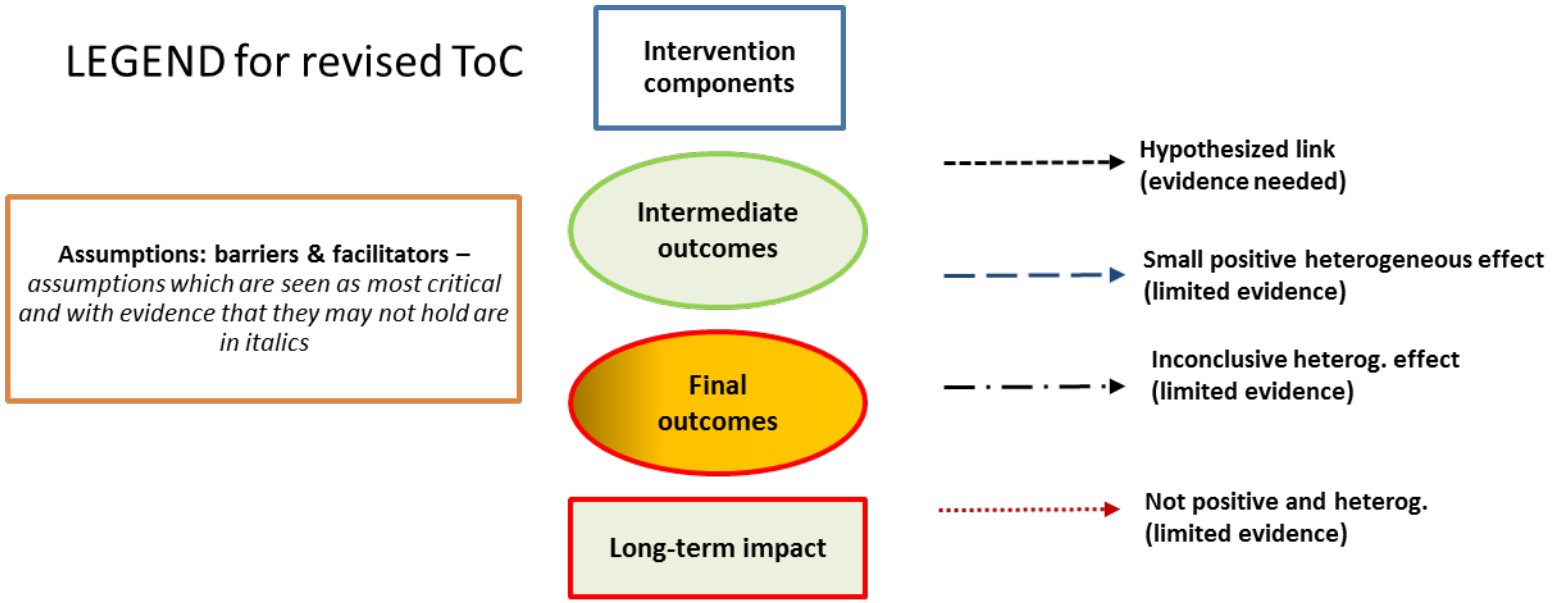
CS, and a myriad of contextual factors, barriers and facilitators but only a limited number of studies available for quantitative effect sizes per outcome. For some intermediate outcomes there is more evidence (yields and farm income in particular), but we find little on more immediate outputs (such as PO strengthening, training and knowledge, and input provision) and final outcomes, with the exception of household income, for which there is a bit of evidence that can be combined with evidence on farm incomes.

The other – unsurprising – conclusion, namely that context matters more than anything, is the most robust result of this exercise. The variety of effect sizes across outcomes and CS makes it impossible to produce a general statement about whether certification works or does not work for the wellbeing of producers and workers. There are differences between CS and differences in the contexts in which each CS works. Although for some outcomes some CS do better than others, it is not advisable to reach a broader conclusion in this respect for three main reasons. First, heterogeneity remains significant even within a single CS, i.e. a CS may do well in some places, or for some outcomes, but not for others. Second, the quantitative evidence synthesized is very limited, given the small number of studies per outcome. Third, the number of CS for which we have ES for different outcomes is also limited, which reflects a bias in research towards some CS, while for many other CS we find no evidence at all. The differences in terms of context and how the bundles of interventions operate on the ground are important and explain a good deal of the heterogeneity found. Thus, all we can say is that there are instances in which certification and their interventions work for some outcomes, but these are limited in number and highly contextual. Equally we find instances in which the expected outcomes are not consistently achieved, such as the case of labour standards.

Although there is substantial evidence extracted from the qualitative synthesis that helps us understand the conditions under which the effectiveness of CS is constrained or enhanced (more of the former as most of the extracted evidence is on barriers), there is very limited evidence from ‘linked studies’, i.e. qualitative insights from quantitative impact evaluations included in meta-analysis. Most studies in this category report very basic, and generally rather superficial, information, that could not be used to provide a deeper analysis of barriers and facilitators for implementation dynamics, distribution and other external contextual factors. Therefore, this synthesis proposed a number of key contextual factors that are known to affect the hypothesised linkages in the revised ToC more systematically and consistently across a range of geographical and institutional contexts. We have mentioned several key assumptions that need to be addressed in each case, which can be linked to the four different groups of interventions identified by the revised ToC.

Figure 39 - Revised theory of change: integrated synthesis of evidence




LEGEND for revised ToC
Intervention components
Intermediate outcomes
Final outcomes
Long-term impact
Assumptions: barriers & facilitators – assumptions which are seen as most critical and with evidence that they may not hold are in italics
Hypothesized link (evidence needed)
Small positive heterogeneous effect (limited evidence)
Inconclusive heterog. effect (limited evidence)
Not positive and heterog. (limited evidence)

5. Discussion

5.1 Summary of main results

This systematic review found 43 studies eligible for a quantitative synthesis of effects of certification schemes in agricultural production in LMICs. We considered evidence on a wide range of intermediate and endpoint outcomes, based on a theory of change that considered various causal chains related to different kinds of interventions, which normally form part of the certification process. The range of interventions and outcomes was wide because, as argued in Section 1, CS operate with bundles of interventions depending on the specific standards required by each CS, and specific interventions actually implemented in each context.

Given the nature of the topic, the broad scope of the review, and the type of literature that has generated evidence on CS, the review included different types of studies, from journal articles and books, to evaluation reports for CS, to unpublished material, including PhD theses and other outputs. This underscores the need for careful hand-searching in systematic reviews of this kind. The implication is important in terms of time and resources, since much more time and many more hours of work are needed when a large proportion of studies need to be searched through hand 'targeted' searching. Moreover, many websites where relevant documents could be found did not allow complex search strings and required a painstaking exercise of trying multiple searches with simple terms. Still, a large number of studies have been identified and used for both the quantitative and qualitative syntheses. It is clear that generally many suffer from high risk of bias, as noted in the section below.

In terms of settings there is no single one that dominates clearly, but many studies report on Latin America and Africa. There is a dearth of evidence on Asia, which does not necessarily mean that the outreach of standards in Asia is less important than in other regions. This pattern in part reflects the predominance of studies on Fairtrade, which is concentrated precisely in Africa and Latin America.

In summary, as suggested in the integrated synthesis, and despite the wealth of information extracted from qualitative studies, we find that the available evidence does not give a clear picture of the impact – or lack thereof – of certification schemes.

We calculated effect sizes for prices, yields, certified farm income, wages, household income, assets, illness and schooling. We have the most evidence for income from certified production. Ten studies return a positive pooled effect (SMD 0.22, 95%-CI from 0.03 to 0.41) rise in income for certified producers. The overall effect is statistically significant (p=0.021). On wages however, across eight studies we find that certification lowers wages of workers in agricultural production (SMD -0.26, 95%-CI from -0.46 to -0.06). This effect is statistically significant (p=0.012). Possibly one of the most important outcomes for farmers is the change in total household income as a result of certification. Here eight studies combine to show an overall increase in total household income as a result from certification (SMD 0.13, 95%-CI from -0.06 to

0.32). The pooled effect is however not statistically significant (p=0.17). The evidence base is weakest for effects on wealth and illness, as we have evidence from just two studies for each. For wealth the joint effect is a slight increase in wealth among certified producers (SMD 0.05, 95%-CI from -0.15 to 0.26), but the effect is not statistically significant (p=0.598). The meta-analysis for illness finds a negative pooled effect (SMD -0.15, 95%-CI from -0.32 to 0.03, i.e. illness is reduced) though again the effect is not statistically significant (p=0.106). Finally, for schooling we estimate an increase in school attendance as a result of certification (SMD 0.12, 95%-CI from 0.00 to 0.24). The pooled effect, which comes from five studies, is statistically significant (p=0.041)[96]. We have subjected these findings to thorough sensitivity analysis, which showed few systematic patterns, and we also have little reason to believe that the findings are compromised by publication bias.

While the evidence in hand points largely towards findings that are not statistically significant, the evidence base is also too thin in most cases to have great confidence in these results. This is in large part the result of the limited number of impact evaluations that met the inclusion criteria for this review. Where we do have substantially more data is on qualitative evidence, looking at barriers and facilitators, and an array of different kinds of contextual factors. The summary table in Section 4.5 of the report (the qualitative synthesis) provides a list of key factors that mediate the linkages between CS interventions and intermediate and final outcomes, organised around the themes of implementation issues, distributional dynamics, and other contextual factors and barriers and facilitators Context matters, but it matters in a wide variety of ways depending on the type of intervention, the type of causal chain and the type of setting.

5.2 Overall completeness and applicability of evidence

We have used 43 studies to address questions of effectiveness for a total of eight outcomes. The qualitative synthesis draws on 138 studies, reporting on a large number of issues that we organized around three main blocks: barriers and facilitators in implementation; distributional dynamics; and other contextual factors and barriers and facilitators.

This review has primarily shown that there is a serious lack of reliable evidence on the effects of CS. We have reviewed evidence for a range of intermediate and final outcomes. There are some outcomes for which we found no evidence at all among included studies, in particular empowerment did not feature in our effectiveness review for this reason.

For those outcomes for which we have more evidence, only one (farm income) had 10 studies for meta-analysis. Considering that some the studies included in that outcome had high RoB rating, we are left with a limited number of studies on which to make any claims about effectiveness. There are several outcomes for which the number of studies is inadequate, especially considering the number of them with high

[96] The pooled effect is however no longer statistically significant if unit of analysis corrections are applied, see Annex F.

RoB ratings: health (illness) has two studies, education (schooling) has five, three of which with high RoB, price only three and yields only five. Considering that for each of these outcomes there are important moderators (commodities, type of programme, RoB) and also significant heterogeneity, the applicability of this evidence is limited.

It is also important to note that these results do not permit us to say much about the effectiveness of most of the individual interventions considered as part of CS. Studies report on the fact of being certified, which is used as a proxy for certain types of interventions, but without enough precision. Therefore, the meta-analysis can say whether Fairtrade certification seems more or less effective in terms of farm or household income for a limited number of studies, but cannot say whether the floor price or the support given to POs is effective or not in terms of these outcomes. We can only learn from insights extracted from qualitative studies into possible reasons or barriers that would make the price intervention more or less effective. This is an important finding which raises the question as to whether impact evaluations should concentrate on individual interventions within one scheme, or continue reporting on the status of certification as a proxy for a bundle of interventions that cannot then be disentangled.

5.3 Quality of the evidence

Generally the quality of the evidence was variable and the proportion of studies with high RoB ratings was relatively high. There were no RCTs and a range of quasi-experimental designs employing different techniques of data analysis. Given the paucity of calculable effect sizes per outcome and the variety of methods the meta-analysis encountered difficulties.

One of the reasons underlying the high and even critical RoB ratings of many studies was the lack of sufficient reporting of – sometimes very basic – aspects of study design, sampling and data collection. Partly this is done to remain user-friendly, though technical annexes can be useful way of separating out technical detail in this case. As it stands, too many studies are unclear on how treatment and control groups are selected and how exactly data were collected. Many fail to present the statistical information necessary to judge the adequacy of their effects of overcome selection bias and confounding factors. Even otherwise well-designed and seemingly well-executed studies can fail to report vital pieces of information. Uneven reporting practices also impact the calculation of effect sizes. Many studies do not report the exact size of the estimated effect, instead indicating only direction and statistical significance of an effect. Others fail to report p-values and standard errors (or confidence intervals). Both practices mean the reader is forced to rely on the judgements presented by study authors regarding the substantive and statistical significance of effects. Where studies fail to report important statistical information such as variances, standard errors, exact p-values, t-values or (in some cases) sample sizes, the calculation of standardized effect sizes becomes difficult or impossible.

The group of studies used for RQ2, most of a qualitative nature and some extracted from the list of quantitative evaluations, also displayed a variety of methods and RoB

ratings. An initial distinction was made between ethnographies and non-ethnographic studies, the former being generally richer in terms of depth of analysis, but also in reporting on methods and issues such as reflexivity and triangulation. Despite the fact that, overall, the group of 138 studies provided a wealth of information on barriers and enablers in implementation dynamics, on distributional forces, and on other external contextual factors, their quality was variable, and several studies, especially non-ethnographic contributions, were only borderline in terms of minimum reporting standards.

In sum, given the prevalence of studies with mixed quality, including a significant number of high RoB studies, the results of the review should be interpreted with much caution. For RQ1 for instance only 12% of studies were rated as being free of selection bias and confounding, while only 5% were rated as having done enough to ensure the equivalence of treatment and control groups. In a sense, the methodological lessons are more reliable than the findings on effects of CS on socio-economic outcomes.

5.4 Limitations and potential biases in the review process

The review established strict criteria for inclusion, based on the need for a comparison group and methods to control for sources of confounders. However, during the data extraction some doubts were raised about a number of studies and potential selection bias or insufficient controls for confounders. A number of assumptions also had to be made to increase the number of effect sizes from studies that failed to report some basic statistical information.

Moderator analysis could only be applied to selected meta-analyses due to the low number of available and usable observations. Generally the measured degree of heterogeneity was quite high in most forest plots.

The searching process was quite comprehensive and even included a number of languages other than English. However, the number of relevant studies in foreign languages meeting the selection criteria was quite limited. Despite the fact that reviewers read material in Spanish, French, Portuguese and German, the final set of studies was almost unanimously in English, suggesting that perhaps some non-English studies may not have been found in normal and targeted searches.

An important limitation of this review is that we have a large number of studies with substantial relevant evidence for key themes identified for RQ2 but very limited evidence on barriers and facilitators from quantitative impact evaluations under RQ1. In other words, it is impossible to provide a deeper analysis of the causal chain for RQ1 studies and therefore be more specific about the conditions of those specific interventions in those contexts. However, a large number of non-linked qualitative studies do offer substantial insights into barriers and facilitators that should be relevant to the contexts of those studies which have produced usable effect sizes.

5.5 Agreements and disagreements with other studies or reviews

As noted in Section 1.4, we did not find relevant reviews of the literature that met the criteria for a systematic reviews. However, a number of less systematic reviews do exist and some of these use systematic methods of searching for relevant studies. In particular these are Nelson and Pound (2009), Blackman & Riviera (2010) and ITC (2011). These reviews depart from systematic review methodologies in their methods of synthesis, which rely on vote counting. The major shortcoming of vote counting is its failure to take systematic account of either the magnitude of effects or the precision with which the reported effects are estimated. None of these reviews presented a meta-analysis of impact estimates. We also searched for systematic reviews on related topics, as these may contain findings relevant to our own review questions. We identified just one such review, namely Ton et al (2016) on contract farming, which is was not yet published in the Campbell Collaboration Library, but for which a paper with preliminary findings was already available at the time of writing. In this section we compare our findings to those presented in these reviews.

Blackman and Rivera (2010)'s review has a wider scope than ours, as it includes literature on the impacts of certification on producers in agriculture, tourism, fisheries and forestry. The review also includes studies looking outcomes in high-income countries. The authors identify 37 empirical studies, of which they judge 14 to be of adequate methodological quality. To be judged of adequate quality a study has to present a robust counterfactual. Similar to our own findings the authors find that a large proportion of studies use matching methods to address selection bias. As in our review Fairtrade is by far the most studied standard among the studies identified by Blackman and Rivera. The results are sorted according to the sectors studied, rather than more precise outcome measures, so we can only compare findings in a very broad sense. Of the 14 studies included in the discussion of impact estimates, only six find that certification has a positive impacts. Blackman and Rivera (2010, p. 15) summarise their findings as follows: 'Hence, at best, the [counterfactual impact] studies provide very weak evidence for the hypothesis that sustainable certification has positive socioeconomic or environmental impacts'. At this level of abstraction the findings are line with our own, which found either modestly positive or statistically insignificant pooled effects for most outcomes.

The literature review published by the ITC (2011), the second in a four-part series on certification schemes, expended considerable effect to undertake and document a systematic search and study selection process. Their review also had a wider scope than our own, including studies on the impact on producers in forestry, fisheries and livestock, in addition to agriculture. The authors identify 47 empirical papers across all sectors, of which they judged 19 to present 'strong counterfactual outcomes' (ITC, 2011, p. 18). The other 28 papers did not present a counterfactual analysis, but were judged to be of sufficient methodological quality to inform questions not directly concerned with impact. Mirroring our own search results, the majority of the papers included by the ITC deal with Fairtrade certification, and a plurality are focused on coffee.

The ITC present their results, which they arrived through vote counting and narrative synthesis, separately for different outcomes, allowing for a more fine-grained comparison to our own results. On yields, they find mixed results with three studies reporting positive impacts, against three reporting no impact or mixed results and three with negative findings. We found a moderately negative pooled effect for yields, though this was not statistically significant and most individual studied found no effect. With regards to prices the evidence collected by the ITC is overwhelmingly positive, which is in line the positive and statistically significant pooled effect we identified. Looking at net incomes after certification, the ITC finds mostly positive results (eight studies), but also four studies reporting null or mixed results and two studies reporting negative impact. This seems compatible with our own result on net income from certified production, which was a moderately positive and statistically significant pooled effect. Our meta-analysis included three studies reporting negative effect, though the effect size estimates for none of these studies was statistically significant. The ITC review also found very positive results for the overall impact on producers' livelihoods. Though it is not clear which measures were used, the closest comparison are our results on total household income. Where the ITC find nine studies reporting positive impacts and three studies reporting null or mixed results, our meta-analysis finds a positive pooled effect, that is however not statistically significant. We include three studies with negative effect size estimates, though none of these are statistically significant. In their summary of their results, the ITC authors report that: 'Overall, the direct impact of participating in private standards in terms of price and profits received by producers tended to be positive, even when compared to alternatives. However, this was not a uniform conclusion. A number of studies also found mixed evidence on the net income for producers and some even found a negative impact on net income for producers, where the increased earnings did not compensate for the additional costs and increased labour involved in complying with standards requisites' (ITC, 2011, p. 23). These findings are supported by our own, only that our positive finding on total household income was not statistically significant.

Nelson and Pound (2009), in a study commissioned by the Fairtrade Foundation, provide a conventional literature review on the impact of Fairtrade certification. So the first difference with our review is the scope, as only one CS is considered. However, our results include a disproportionate number of studies reporting on Fairtrade (52% overall) and an absence of studies reporting on a large number of relevant CS with social sustainability standards. Therefore, a comparison with Nelson and Pound (2009) is relevant. Nelson and Pound (2009) reviewed 23 papers with 33 case studies, from an initial stock of over 80 publications. It is possible that they missed some studies as this review has found a large number of items reporting on Fairtrade before and after strict inclusion criteria were applied (128 studies for RQ1 and RQ2 combined after screening). It is true that, as shown in Section 4.1, the number of rigorous quantitative impact evaluations has accelerated since 2009, so Nelson and Pound (2009) covered the literature before this expansion. There are clear similarities in the descriptive results about the characteristics of the literature on Fairtrade, especially the areas of focus and main research gaps. So, they find a major focus of reviewed studies on Latin America, coffee, and small producers. Only

two studies on wage workers in plantations were reviewed, which underscores the dearth of evidence on labour market outcomes in the literature on certification - a gap that has been only very partially addressed in recent years. Our review also presents this skewed coverage in relation to studies of Fairtrade, but the proportion of studies in Nelson and Pound (2009) that focus on coffee and Latin America is even greater than in our case, and our review includes a higher proportion of studies reporting on labour conditions.

In terms of effectiveness, Nelson and Pound (2009) highlight a number of positive effects, which they arrive at by counting the number of papers that mention benefits from specific interventions associated with Fairtrade certification. In many cases these were perceived benefits or positive effects estimated without a counterfactual or adequate control for confounders. It is hard to compare these conclusions with our effort to statistically synthesize calculable effect sizes from rigorous impact evaluations. It is interesting to note that Nelson and Pound (2009) find the literature patchy in terms of drawing general conclusions and findings too context-specific. They conclude that 'further research is needed to establish what are the key factors driving success, as current studies are weak on teasing these out' (2009, p. 5). Most of the report focuses on outcomes and less on barriers and facilitators, despite the fact that many of the studies they used are included in RQ2 of this review and contribute to substantial insights into barriers and facilitators (especially the barriers). Finally, their review finds very limited evidence on final outcomes, such as income, expenditure or assets for participating households, a problem we have also encountered, whereas they find many studies with evidence on whether producers are getting higher prices for their products and improved access to credit, outcomes that are closer to interventions in the causal chain.

Finally, the systematic review of contract farming by Ton et al (2016) is worth noting briefly for two reasons. First, it is a systematic review that follows comparable standards and protocols to this review. Second, contract farming is a contractual arrangement in agriculture that also characterizes, and often overlaps with, the practice of certification. The outcomes of interest are similar (yields, farm incomes) and some of the interventions (price agreements, input access, credit, access to export markets) are similar. Ton et al (2016) reviewed 22 quantitative studies with a strong counterfactual design, which is somewhat surprising given that the literature on contract farming is vast and more generic than that on CS. Their meta-analysis only focused on income effects because of data availability and, unlike our review, they reach high statistical conclusion validity, generally with positive effects, albeit with low construct validity, because they pool effects from very different cases. Like in this review, the heterogeneity is very significant. It is also remarkable that they find a strong publication and survivor bias, 'that are inherent to the area of research', a point that could be potentially applicable to research on CS too, even though we do not find the same results for publication bias. A very important finding, which confirms what other conventional literature reviews had found, is the enormous diversity of contract farming arrangements, with a substantial number of factors mediating the impact of the contractual arrangement on incomes, as is the case with certification. Finally, another interesting similarity is the additional costs and risks

incurred by a producer in joining a contract farming scheme, and the associated segmentation in favour of more prosperous and larger farmers, reflecting the existence of barriers to entry to successful contract farming schemes as is the case with many CS.

6. Authors' conclusions

6.1 Implications for practice and policy

The review assesses the effectiveness of CS for agricultural production in terms of key socio-economic outcomes for agricultural producers and workers, i.e. those who are usually at the bottom of agricultural supply chains. The background section documented the growth of CS and their associated social sustainability standards which this review focuses on. The outreach both in terms of numbers of participants and settings (countries and types of production) has expanded significantly in the past 30 years. However, there is also evidence that the volume of output channeled through some of the best known CS, such as Fairtrade, remains limited in comparison to 'conventional markets'. Therefore, there is still much scope for growth and expansion into new commodities and production systems.

For this reason it is important to assess the evidence on impact of these schemes on the wellbeing of main beneficiaries, i.e. agricultural producers, especially small farmers, and agricultural wage workers. The findings suggest that CS operate with bundles of interventions whose effects on socio-economic outcomes are hard to disentangle. Studies generally focus on the status of producers or workers, in the sense of being certified or not, not on whether they received a premium or a particular training pr0gramme. Therefore, results can only be interpreted in terms of whether CS as bundles of interventions, rather than their individual interventions (such as the price premium), have any impact on key outcomes such as producer prices, yields, farm income, profits, labour outcomes (wages, security of employment, non-wage benefits, etc.), household income, assets and health and education outcomes. The review only found a limited number of studies for each of these outcomes. Therefore the meta-analysis was restricted by an insufficient number of effect sizes per outcome to reach any robust conclusion. Despite this serious limitation, we have extracted 53 separate effect sizes from 29 different studies. We have conducted meta-analyses for each outcome for which we have at least two effect sizes. In total 44 effect sizes were used for 8 main meta-analyses. The main synthesised effects can be summarised as follows:

1. Pooled effects on yields (productivity) are non-significant, with many individual effect sizes close to zero (central estimate -20%, range from -52% to 19%; SMD -0.42, 95%-CI from -1.23 to 0.39).
2. Pooled effects on final producer price are positive and statistically significant but lose significance if we exclude one study with high risk of bias (central estimate 14%, range from 4% to 24%; SMD 0.28, 95%-CI from 0.09 to 0.49).
3. The synthesized effect on certified farm income is a modest and statistically significant positive effect, mostly driven by studies reporting on GlobalGAP and Utz (central estimate 11%, range from 2% to 20%; SMD 0.22, 95%-CI from 0.03 to 0.41).

4. For wages, results suggest that certification is associated with lower wages, a combination of negative effect sizes from a number of studies pertaining to the same project, and zero and non-significant effects for other studies (central estimate -13%, range from -22% to -3%; SMD -0.26, 95%-CI from -0.46 to -0.06).
5. The result for total household income is also a pooled effect that is not statistically significant, with remarkable variation from moderate negative to positive effect sizes even for the same commodity (coffee) (central estimate 6%, range from -3% to 16%; SMD 0.13, 95%-CI from-0.06 to 0.32).
6. Health and assets only have two studies each and the pooled effect is not sufficiently meaningful for interpretation (Assets: central estimate 3%, range from -7% to 13%; SMD 0.05, 95%-CI from -0.15 to 0.26. Health: central estimate -7%, range from -16% to 2%; SMD -0.15, 95%-CI from -0.32 to 0.03).[97]
7. Pooled effects on schooling are positive and statistically significant but three out of five studies find no significant effect (central estimate 6%, range from 0% to 12%; SMD 0.12, 95%-CI from 0.01 to 0.24).

So, overall, the picture is one of mixed results and a dominance of weak or statistically non-significant effects. There are differences between CS, but heterogeneity is also present among studies reporting on the same CS. In other words, for every CS there is very substantial variation. So, it is hard to conclude anything about whether any particular CS performs better compared to others over a range of outcomes. Indeed we find effects that are contradictory for the same outcomes and CS across studies. The degree of heterogeneity of results for most of the meta-analyses performed is very high. Does this mean that CS do not achieve what they set out to achieve? The evidence found is too limited to reach such a conclusion. However, it does raise questions about the multiplicity of sources of variation. CS operate in environments with multiple interventions, goals, actors and contexts. They do not operate in a social, institutional and economic vacuum. Indeed, there is a tendency to locally-specific results, which affect the external validity of impact evaluations. The heterogeneity is such that finding a positive effect in one or more contexts cannot be interpreted as the intervention (certification) generally working for the expected outcomes.

It is probably not surprising that the importance of context and the wide array of possible barriers are key insights from the qualitative synthesis. There are lessons for practice and policy emerging from the analysis of barriers in particular. For example, the causal chain between the CS interventions and outcomes for farmers and workers is held together by a series of nodes along the chain, such as:

- POs and their characteristics, particularly heterogeneity and power relations within them;
- relations with buyers and exporters;
- business models linking buyers and producers (whether open spot markets,

[97] Please note that, as these findings concern illness, a negative synthesised effect means an improvement in health.

contract farming or a mix);
- national institutions shaping the dynamics of agricultural trade and labour relations;
- direct and indirect certification costs, which negatively affect adoption or the size of benefits accruing to producers;
- availability of additional external support, often critical for adoption and sustained maintenance of standards;
- heterogeneity of participant groups and effects of inequality on POs management and sharing of benefits;
- difficulties in addressing deep-rooted structures of inequality based on gender.

These various nodes and their interactions (they never happen in isolation) may contribute to an explanation of the wide variation of effects for intermediate and endpoint outcomes. The array of assumptions usually held to support the hypothesized causal chain underpinning the ToC of CS is substantial. These assumptions need to be considered carefully and explored in each context in order to understand the likelihood or not of success.

Acting on some of the barriers mentioned above could mean a revision of standards, which many CS routinely undertake. We have documented the tendency towards proliferation of standards and growing overlaps between CS in terms of what they require and cover. Perhaps a lesson is that CS could specialize more in specific niches of sustainability, reduce the number of standards and requirements per standard, and tighten monitoring and auditing to focus more on what is achievable. This might lead to less complex bundles of interventions and make evaluations also more meaningful. It could also result in lower certification costs, an issue that has been analysed in this review as a frequently mentioned barrier. In this regard, it would make sense for CS to consider the relative value added of the different interventions they usually ‘bundle’, and be more selective.

Another important recommendation is that CS could perhaps take a careful look at how claims are made, especially in the context of advocacy campaigns to support social sustainability standards and practices of fair trade. Many studies tend to highlight the mismatch between the expectations raised and the claims made by many CS and the participants’ lived experiences, particularly in terms of monetary benefits, but also working conditions and female participation. By making strong claims about long-term impacts, such as poverty reduction, or sustainable farming or empowered producers and workers, CS may be introducing unnecessary risks to their credibility. Therefore, a possible option for CS is to revise their results frameworks and focus more squarely on a more limited set of achievable results that can be made as context-specific as possible. So, if a CS is unlikely to work with certain types of farmers (for example, very poor and small producers) or in supply chains where the potential demand for certified products is constrained, perhaps the outreach and focus of the CS could be reconsidered.

6.2 Implications for research

Probably the most robust conclusion of this review is the need for additional, as well as better, research on effects and their barriers and facilitators, so that a more consistent picture of the causal chains between different types of interventions under different CS and key intermediate and endpoint outcomes can be drawn. This review searched for evidence on multiple certifications, with many different interventions, on a wide range of crops, in LMICs and with two broadly defined groups of participants: agricultural producers and workers. Despite the existence of so many different standards and CS the available literature is skewed towards a certain group of well-known CS. It is striking the extent to which impact evaluations of Fairtrade certifications dominate the literature. Therefore, in order to build a more complete understanding of different causal chains for different types of CS, more research is needed on the standards and schemes that are least researched. Even the evidence on effects of Fairtrade certification is far from conclusive, at least from the point of view of quantitative effects. Despite the large volume of literature on this scheme there is a dearth of high-quality quantitative impact evaluations for key outcomes such as producers' and workers' incomes, and especially health and education outcomes.

The other key message that arises from this review is methodological. There are two main considerations. The requirements for high-quality impact evaluations are demanding. Up until now, only the last four or five years have seen a substantial increase in the number of evaluations that use methods that are usually associated with adequate control for confounding and selection bias. We did not find any studies that met the inclusion criteria for RQ1 published before 2008. Nonetheless, much research had already been conducted on CS by then, and important claims about impact made in different ways. Therefore, both impact evaluation departments of CS and independent researchers need to catch up with the methodological demands for high-quality research in this field, and understand the methodological and logistical challenges that they may entail when conducting primary research. The second point is that given the nature of the intervention, the stakeholders involved and the modus operandi of CS conducting high-quality experimental evaluations is likely to be a challenge, even if not impossible. The costs of conducting high-quality quantitative impact evaluations (both experimental and quasi-experimental) are substantial, especially if follow-up surveys are necessary and the reporting frequency is high. A clearly defined intervention would need to be the focus, rather than treating the certification status as 'intervention'. This can be achieved through theory-based approaches but would have to focus on selected component rather than on the scheme as a whole. It is unlikely that a CS would restrict its domain of action to a single intervention so that an RCT can be operationalized, but this cannot be discarded especially if CS become increasingly conscious of the need to estimate the impact of specific interventions. One RCTs looking at the effects of certification that is currently underway faces a similar issue. In the baseline report of their study on coffee certification in Indonesia Neilson & Toth (2016) note that finding a clear baseline had proved difficult and that the certification was often implemented at the same time as other support programmes, making it difficult to disentangle effects.

The authors therefore decided to compare a 4C-certified control group with another 4C-certified group, which was in the process of becoming RA certified. An ‘untreated’ control was available in the context of the programme. Furthermore, since most CS operate at level of collective organisations (POs) an RCT would need to be implemented at grouped level, although effects need to be analysed at individual producer level. Selection bias can happen at cluster and at individual level at the same time. Finally, unlike post-hoc evaluations, experimental designs require CS to tailor interventions to the basic need of randomization. For instance, in their ongoing study on BCI certification in India a team of researchers was able to work with BCI to randomise an already planned rollout of the CS so as to create the conditions for an RCT comparing certified and non-certified groups (Kumar et al, 2015). While this may be desirable for evaluation purposes, it may not always make sense for business or operational reasons from the point of view of the CS though.

The point, and our suggestion, in light of the importance of context, is that researchers aiming to shed more light on causal chains for CS interventions should opt for theory-based mixed-methods evaluations, with a strong feasible quasi-experimental component, making sure that more evidence on implementation and process is adequately collected in order to link effect results with evidence on barriers and facilitators. There is already scope for improvement in current impact evaluations even without experimental methods. There is also scope for improvement in econometric analysis if more detailed evidence on variability of implementation dynamics and processes is collected, coding the different configurations of intervention components for analytical purposes. Given the importance of the costs of certification for producers, it may also be advisable to complement such impact evaluations with cost-effectiveness analysis, especially for those CS that are more expensive and require important investments from producers.

The final suggestion refers to reporting protocols. One problem faced in the process of this review was finding relevant information in the right places. It starts with abstracts, which sometimes do not even give indication that the study is based on secondary sources or entails data collection, or what outcomes of interest are studied. Then, for both quantitative and qualitative studies, the amount of detail on methods used tends to be limited, and often insufficient to meet some of the inclusion criteria usually applied in systematic reviews, or even to arrive at a fair judgements about the methodological merits of the study. This is of course a general problem in research on international development, but we found an excessively large number of studies that could potentially be useful, but did not report enough methodological information on both data collection and data analysis to be included in the review. Sometimes this reflects biases in certain publication outlets, which prefer authors to focus on findings and leave technical and methodological detail aside. Unfortunately, the only way we can assess the foundations of research findings is by having enough information on methods used. Authors should be encouraged to consult different options for risk of bias tools in order to anticipate possible problems of bias and correct designs accordingly. The need for better reporting of methods and details on analysis and tests does not only concern quantitative impact evaluations. Qualitative

studies should follow the example and report more on critical issues such as the justification for research site selection, detailed descriptions of context, some information on how respondents were selected, the influence of the researcher's position, the triangulation undertaken and any question about external validity that may be worth considering in order to assess the wider implications of qualitative findings. Ethnographies tend to do well in this respect, so rapid appraisals based on qualitative methods should strive to adopt these standards of reporting.

Annex A: Search strategy examples

Web of Science-SSCI (Searched 2nd May 2015)

<1990 to 2015 Week 18 >

16

1,498

#15 AND #7 AND #4

Indexes=SSCI Timespan=1990-2015

15

370,289

#14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8

Indexes=SSCI Timespan=1990-2015

14

6,607

TS=((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income") NEAR (economy or economies))

Indexes=SSCI Timespan=1990-2015

13

187

TS="transitional countr*"

Indexes=SSCI Timespan=1990-2015

12

2,607

TS=(low NEAR/3 middle NEAR/3 countr*)

Indexes=SSCI Timespan=1990-2015

11

4,039

TS=(lmic or lmics or "third world" or lamicountr*)

Indexes=SSCI Timespan=1990-2015

10

654

TS=(low NEAR (gdp or gnp or "gross domestic" or "gross national" or GNI))

Indexes=SSCI Timespan=1990-2015

9

66,304

TS=((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income" or underserved or "under served" or deprived or poor*) NEAR (countr* or nation? or population? or world or economy or economies))

Indexes=SSCI Timespan=1990-2015

8

326,981

TS=(Afghanistan or Angola or Albania or "American Samoa" or Argentina or Armenia or Armenian or Azerbaijan or Bangladesh or Belarus or Belize or Benin or Bolivia or Bosnia or Herzegovina or Botswana or Brazil or Bulgaria or Burkina Faso or Burkina Fasso or Burundi or Urundi or Cambodia or Cameroon or Cameroons or Cameron or Camerons or Central African Republic or Chad or Chile or China or Colombia or Comoros or Comoro Islands or Comores or Congo or Costa Rica or Cuba or Zaire or Cote d'Ivoire or Ivory Coast or Djibouti or Dominica* or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Ethiopia or Fiji or Gabon or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or Hungary or India or Indonesia or Iran or Iraq or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizstan or Lao PDR or Laos or Lebanon or Lesotho or Liberia or Libya or Macedonia or Madagascar or Malagasy Republic or Malawi or Malaysia or Maldives or Marshall Islands or Mali or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or Nicaragua or Niger or Nigeria or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or Romania or Rwanda or Ruanda or Samoa or Samoan Islands or Sao Tome or Senegal or Serbia or Seychelles or Sierra Leone or Sri Lanka or Solomon Islands or Somalia or South Africa or St Lucia or St Vincent or Grenadines or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Tonga or Togo or Togolese Republic or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or Uruguay or Uzbekistan

or Uzbek or Vanuatu or Venezuela or New Hebrides or Vietnam or Viet Nam or West Bank or Yemen or Zambia or Zimbabwe)

Indexes=SSCI Timespan=1990-2015

7

210,325

#6 OR #5

Indexes=SSCI Timespan=1990-2015

6

70,400

TS=(coffee OR cocoa OR tea OR infusion* OR "yerba mate" OR "camomile" OR sugar* OR fruit* OR banana* OR pineapple* OR mango* OR coconut* OR apricot* OR nut* OR cashew* OR "shea butter" OR argan OR rice OR quinoa OR bean* OR chickpea* OR "red kidney" OR lentil* OR soy* OR herb* OR spice* OR "olive oil" OR olive* OR wine OR honey OR cotton OR flower* OR floriculture OR "palm oil" OR (crop* NEAR/2 produc*))

Indexes=SSCI Timespan=1990-2015

5

149,299

TS=(Farmer* or farming or agricultur* or horticultur* or grower* or producer* or worker* or labo?rer* or smallholder* or small-holder* or cooperative* or co-operative* or syndicate* or ((trade or labo?r) NEAR union*) or "agricultural sector" or "agricultural trade" or "floriculture" or "crop production" or "agricultural products")

Indexes=SSCI Timespan=1990-2015

4

23,229

#3 OR #2 OR #1

Indexes=SSCI Timespan=1990-2015

3

2,207

TS=("fair trade" or fairtrade or fair-trade or transfair or "fair for life" or "Rainforest Alliance" or "Sustainable Agriculture Network" or "UTZ Certified" or "UTZ" or "Global Partnership for Good Agricultural Practice" or "Global GAP" or "GlobalGAP" or "4C

Association" or "Better Cotton Initiative" or "BCI" or "Cotton made in Africa" or Bonsucro or "Ethical Tea Partnership" or Trustea or "International Federation of Organic Agriculture Movements" or IFOAM or "soil association" or "IOAS" or "LEAF" or "Linking Environment and Farming" or "Union for Ethical BioTrade" or "UEBT" or "Roundtable on Sustainable Palm Oil" or "RSPO" "Fair Flowers Fair Plants" or "ProTerra" or "ISO 14001")

Indexes=SSCI Timespan=1990-2015

2

16,368

TS=((fair* OR ethic* OR alternative OR sustainab* OR responsib* OR specialty OR eco OR ecologic OR ecological OR organic) NEAR/3 (certifi* OR standard* OR label* OR seal* OR scheme* OR trad* OR market* OR "value chain*" OR commodit* OR product*))

Indexes=SSCI Timespan=1990-2015

1

6,717

TS=("certification" or "quality standards" or "quality label?ing" or "sustainability standards")

Database: CAB Abstracts (Searched 5th May 2015)

<1990 to 2015 Week 17 >

Search Strategy:

---

1 ("certification" or "quality standards" or "quality labelling").sh. (8809)

2 ((fair* or ethic* or alternative or sustainab* or responsib* or specialty or eco or ecologic or ecological or organic) adj3 (certifi* or standard* or label* or seal* or scheme* or trad* or market* or "value chain*" or commodit* or product*)).ti,ab. (38092)

3 ("fair trade" or fairtrade or fair-trade or transfair or "fair for life" or "Rainforest Alliance" or "Sustainable Agriculture Network" or "UTZ Certified" or "UTZ" or "Global Partnership for Good Agricultural Practice" or "Global GAP" or "GlobalGAP" or "4C Association" or "Better Cotton Initiative" or "BCI" or "Cotton made in Africa" or Bonsucro or "Ethical Tea Partnership" or Trustea or "International Federation of Organic Agriculture Movements" or IFOAM or "soil association" or "IOAS" or "Linking Environment and Farming" or "Union for Ethical BioTrade" or "UEBT" or "Roundtable

on Sustainable Palm Oil" or "RSPO Fair Flowers Fair Plants" or "ProTerra" or "ISO 14001").ti,ab. (1566)

4 or/1-3 (46171)

5 (Farmer* or farming or agricultur* or horticultur* or grower* or producer* or worker* or labo?rer* or smallholder* or small-holder* or cooperative* or co-operative* or syndicate* or ((trade or labo?r) adj union*) or "agricultural sector" or "agricultural trade" or "floriculture" or "crop production" or "agricultural products").ti,ab. (527410)

6 (coffee or cocoa or tea or infusion* or "yerba mate" or "camomile" or sugar* or fruit* or banana* or pineapple* or mango* or coconut* or apricot* or nut* or cashew* or "shea butter" or argan or rice or quinoa or bean* or chickpea* or "red kidney" or lentil* or soy* or herb* or spice* or "olive oil" or olive* or wine or honey or cotton or flower* or floriculture or "palm oil" or (crop* adj2 produc*)).ti,ab. (1317998)

7 or/5-6 (1678981)

8 (Afghanistan or Angola or Albania or "American Samoa" or Argentina or Armenia or Armenian or Azerbaijan or Bangladesh or Belarus or Belize or Benin or Bolivia or Bosnia or Herzegovina or Botswana or Brazil or Bulgaria or Burkina Faso or Burkina Fasso or Burundi or Urundi or Cambodia or Cameroon or Cameroons or Cameron or Camerons or Central African Republic or Chad or China or Colombia or Comoros or Comoro Islands or Comores or Congo or Costa Rica or Cuba or Zaire or Cote d'Ivoire or Ivory Coast or Djibouti or Dominica* or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Ethiopia or Fiji or Gabon or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or Hungary or India or Indonesia or Iran or Iraq or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizstan or Lao PDR or Laos or Lebanon or Lesotho or Liberia or Libya or Macedonia or Madagascar or Malagasy Republic or Malawi or Malaysia or Maldives or Marshall Islands or Mali or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or Nicaragua or Niger or Nigeria or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or Romania or Rwanda or Ruanda or Samoa or Samoan Islands or Sao Tome or Senegal or Serbia or Seychelles or Sierra Leone or Sri Lanka or Solomon Islands or Somalia or South Africa or St Lucia or St Vincent or Grenadines or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Tonga or Togo or Togolese Republic or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or Uzbekistan or Uzbek or Vanuatu or Venezuela or New Hebrides or Vietnam or Viet Nam or West Bank or Yemen or Zambia or Zimbabwe).hw,ti,ab,cp. (1934477)

9 ((developing or less* developed or under developed or underdeveloped or middle income or low* income or underserved or under served or deprived or poor*) adj (countr* or nation? or population? or world)).ti,ab. (42827)

10 ((developing or less* developed or under developed or underdeveloped or middle income or low* income) adj (economy or economies)).ti,ab. (688)

11 (low* adj (gdp or gnp or gross domestic or gross national)).ti,ab. (41)

12 (low adj3 middle adj3 countr*).ti,ab. (1656)

13 (lmic or lmics or third world or lami countr*).ti,ab. (2218)

14 transitional countr*.ti,ab. (75)

15 exp developing countries/ (1301383)

16 or/8-15 (1997188)

17 4 and 7 and 16 (11941)

18 limit 17 to yr="1990 -Current" (11857)

19 limit 18 to (english or french or german or spanish) (9941)

20 social sciences/ or contracts/ or franchises/ or regulations/ (22225) – **Broad subject headings**

21 exp sustainability/ (62241) - **subject heading**

22 ("aa000" or "dd100" or "dd500").xc. or "uu000".cc. or "uu450".xc. or "uu460".xc. or "uu470".xc. or "cc000".xc. or "ee110".cc. (365657) **(Classification Codes: Agriculture (General); Agencies & organizations; Laws & regulations; Community participation & development; Community development; Participation & self-help; Social sciences (general); Agricultural economics)**

23 or/20-22 (408200)

24 19 and 23 (5335)

Database: Econlit (EBSCO) (Searched 6th May 2015)

Limiters - Published Date: 19900101-20150531

S19 Search Results: 3,271

S5 AND S9 AND S18

S18

S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17

S17

TI ("transitional countr*") OR AB ("transitional countr*") OR SU ("transitional countr*")

S16

TI ((lmic or lmics or "third world" or "lami countr*")) OR AB ((lmic or lmics or "third world" or "lami countr*")) OR SU ((lmic or lmics or "third world" or "lami countr*"))

S15

TI (low N3 middle N3 countr*) OR AB (low N3 middle N3 countr*) OR SU (low N3 middle N3 countr*)

S14

TI (((low* N1 (gdp or gnp or "gross domestic" or "gross national" or GNI))) OR AB (((low* N1 (gdp or gnp or "gross domestic" or "gross national" or GNI))) OR SU (((low* N1 (gdp or gnp or "gross domestic" or "gross national" or GNI)))

S13

TI (((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income") N1 (economy or economies))) OR AB (((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income") N1 (economy or economies))) OR SU (((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income") N1 (economy or economies)))

S12

TI (((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income" or underserved or "under served" or deprived or poor*) N1 (countr* or nation* or population* or world))) OR AB (((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income" or underserved or "under served" or deprived or poor*) N1 (countr* or nation* or population* or world))) OR SU (((developing or "less* developed" or "under developed" or underdeveloped or "middle income" or "low* income" or underserved or "under served" or deprived or poor*) N1 (countr* or nation* or population* or world))) Show Less

S11

TI ((Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America")) OR AB ((Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America")) OR SU ((Africa or Asia or Caribbean or "West Indies" or "South America" or "Latin America" or "Central America"))

S10

TI ((Afghanistan or Angola or Albania or "American Samoa" or Argentina or Armenia or Armenian or Azerbaijan or Bangladesh or Belarus or Belize or Benin or Bolivia or Bosnia or Herzegovina or Botswana or Brazil or Bulgaria or Burkina Faso or Burkina Fasso or Burundi or Urundi or Cambodia or Cameroon or Cameroons or Cameron or Camerons or Central African Republic or Chad or Chile or China or Colombia or Comoros or Comoro Islands or Comores or Congo or Costa Rica or Cuba or Zaire or Cote d'Ivoire or Ivory Coast or Djibouti or Dominica* or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Ethiopia or Fiji or Gabon or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or Hungary or India or Indonesia or Iran or Iraq or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizstan or Lao PDR or Laos or Lebanon or Lesotho or Liberia or Libya or Macedonia or Madagascar or Malagasy Republic or Malawi or Malaysia or Maldives or Marshall Islands or Mali or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or Nicaragua or Niger or Nigeria or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or Romania or Rwanda or Ruanda or Samoa or Samoan Islands or Sao Tome or Senegal or Serbia or Seychelles or Sierra Leone or Sri Lanka or Solomon Islands or Somalia or South Africa or St Lucia or St Vincent or Grenadines or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Tonga or Togo or Togolese Republic or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or Uruguay or Uzbekistan or Uzbek or Vanuatu or Venezuela or New Hebrides or Vietnam or Viet Nam or West Bank or Yemen or Zambia or Zimbabwe)) OR AB ((Afghanistan or Angola or Albania or "American Samoa" or Argentina or Armenia or Armenian or Azerbaijan or Bangladesh or Belarus or Belize or Benin or Bolivia or Bosnia or Herzegovina or Botswana or Brazil or Bulgaria or Burkina Faso or Burkina Fasso or Burundi or Urundi or Cambodia or Cameroon or Cameroons or Cameron or Camerons or Central African Republic or Chad or Chile or China or Colombia or Comoros or Comoro Islands or Comores or Congo or Costa Rica or Cuba or Zaire or Cote d'Ivoire or Ivory Coast or Djibouti or Dominica* or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Ethiopia or Fiji or Gabon or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or Hungary or India or Indonesia or Iran or Iraq or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizstan or Lao PDR or Laos or Lebanon or Lesotho or Liberia or Libya or Macedonia or Madagascar or Malagasy Republic or Malawi or Malaysia or Maldives or Marshall Islands or Mali or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or Nicaragua or Niger or Nigeria or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or

Phillippines or Romania or Rwanda or Ruanda or Samoa or Samoan Islands or Sao Tome or Senegal or Serbia or Seychelles or Sierra Leone or Sri Lanka or Solomon Islands or Somalia or South Africa or St Lucia or St Vincent or Grenadines or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Tonga or Togo or Togolese Republic or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or Uruguay or Uzbekistan or Uzbek or Vanuatu or Venezuela or New Hebrides or Vietnam or Viet Nam or West Bank or Yemen or Zambia or Zimbabwe)) OR SU ((Afghanistan or Angola or Albania or "American Samoa" or Argentina or Armenia or Armenian or Azerbaijan or Bangladesh or Belarus or Belize or Benin or Bolivia or Bosnia or Herzegovina or Botswana or Brazil or Bulgaria or Burkina Faso or Burkina Fasso or Burundi or Urundi or Cambodia or Cameroon or Cameroons or Cameron or Camerons or Central African Republic or Chad or Chile or China or Colombia or Comoros or Comoro Islands or Comores or Congo or Costa Rica or Cuba or Zaire or Cote d'Ivoire or Ivory Coast or Djibouti or Dominica* or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Ethiopia or Fiji or Gabon or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or Hungary or India or Indonesia or Iran or Iraq or Kazakhstan or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizstan or Lao PDR or Laos or Lebanon or Lesotho or Liberia or Libya or Macedonia or Madagascar or Malagasy Republic or Malawi or Malaysia or Maldives or Marshall Islands or Mali or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Moldova or Moldovia or Moldovian or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or Nicaragua or Niger or Nigeria or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or Romania or Rwanda or Ruanda or Samoa or Samoan Islands or Sao Tome or Senegal or Serbia or Seychelles or Sierra Leone or Sri Lanka or Solomon Islands or Somalia or South Africa or St Lucia or St Vincent or Grenadines or Sudan or Suriname or Swaziland or Syria or Tajikistan or Tadzhikistan or Tadjikistan or Tadzhik or Tanzania or Thailand or Tonga or Togo or Togolese Republic or Tunisia or Turkey or Turkmenistan or Tuvalu or Uganda or Ukraine or Uruguay or Uzbekistan or Uzbek or Vanuatu or Venezuela or New Hebrides or Vietnam or Viet Nam or West Bank or Yemen or Zambia or Zimbabwe)) Show Less

S9

S6 OR S7 OR S8

S8

CC P13 or J54 or P32 or Q13

S7

TI ((coffee OR cocoa OR tea OR infusion* OR "yerba mate" OR "camomile" OR sugar* OR fruit* OR banana* OR pineapple* OR mango* OR coconut* OR apricot*

OR nut* OR cashew* OR "shea butter" OR argan OR rice OR quinoa OR bean* OR chickpea* OR "red kidney" OR lentil* OR soy* OR herb* OR spice* OR "olive oil" OR olive* OR wine OR honey OR cotton OR flower* OR floriculture OR "palm oil" OR (crop* N2 produc*))) OR AB ((coffee OR cocoa OR tea OR infusion* OR "yerba mate" OR "camomile" OR sugar* OR fruit* OR banana* OR pineapple* OR mango* OR coconut* OR apricot* OR nut* OR cashew* OR "shea butter" OR argan OR rice OR quinoa OR bean* OR chickpea* OR "red kidney" OR lentil* OR soy* OR herb* OR spice* OR "olive oil" OR olive* OR wine OR honey OR cotton OR flower* OR floriculture OR "palm oil" OR (crop* N2 produc*))) OR SU ((coffee OR cocoa OR tea OR infusion* OR "yerba mate" OR "camomile" OR sugar* OR fruit* OR banana* OR pineapple* OR mango* OR coconut* OR apricot* OR nut* OR cashew* OR "shea butter" OR argan OR rice OR quinoa OR bean* OR chickpea* OR "red kidney" OR lentil* OR soy* OR herb* OR spice* OR "olive oil" OR olive* OR wine OR honey OR cotton OR flower* OR floriculture OR "palm oil" OR (crop* N2 produc*))) Show Less

S6

TI ((Farmer* or farming or agricultur* or horticultur* or grower* or producer* or worker* or labo?rer* or smallholder* or small-holder* or cooperative* or co-operative* or syndicate* or ((trade or labo?r) NEAR union*) or "agricultural sector" or "agricultural trade" or "floriculture" or "crop production" or "agricultural products")) OR AB ((Farmer* or farming or agricultur* or horticultur* or grower* or producer* or worker* or labo?rer* or smallholder* or small-holder* or cooperative* or co-operative* or syndicate* or ((trade or labo?r) NEAR union*) or "agricultural sector" or "agricultural trade" or "floriculture" or "crop production" or "agricultural products")) OR SU ((Farmer* or farming or agricultur* or horticultur* or grower* or producer* or worker* or labo?rer* or smallholder* or small-holder* or cooperative* or co-operative* or syndicate* or ((trade or labo?r) NEAR union*) or "agricultural sector" or "agricultural trade" or "floriculture" or "crop production" or "agricultural products")) Show Less

S5

S1 OR S2 OR S3 OR S4

S4

TI (("fair trade" or fairtrade or fair-trade or transfair or "fair for life" or "Rainforest Alliance" or "Sustainable Agriculture Network" or "UTZ Certified" or "UTZ" or "Global Partnership for Good Agricultural Practice" or "Global GAP" or "GlobalGAP" or "4C Association" or "Better Cotton Initiative" or "BCI" or "Cotton made in Africa" or Bonsucro or "Ethical Tea Partnership" or Trustea or "International Federation of Organic Agriculture Movements" or IFOAM or "soil association" or "IOAS" or "LEAF" or "Linking Environment and Farming" or "Union for Ethical BioTrade" or "UEBT" or "Roundtable on Sustainable Palm Oil" or "RSPO" "Fair Flowers Fair Plants" or "ProTerra" or "ISO 14001")) OR AB (("fair trade" or fairtrade or fair-trade or transfair or "fair for life" or "Rainforest Alliance" or "Sustainable Agriculture Network" or "UTZ Certified" or "UTZ" or "Global Partnership for Good Agricultural Practice" or "Global

GAP" or "GlobalGAP" or "4C Association" or "Better Cotton Initiative" or "BCI" or "Cotton made in Africa" or Bonsucro or "Ethical Tea Partnership" or Trustea or "International Federation of Organic Agriculture Movements" or IFOAM or "soil association" or "IOAS" or "LEAF" or "Linking Environment and Farming" or "Union for Ethical BioTrade" or "UEBT" or "Roundtable on Sustainable Palm Oil" or "RSPO" "Fair Flowers Fair Plants" or "ProTerra" or "ISO 14001")) OR SU (("fair trade" or fairtrade or fair-trade or transfair or "fair for life" or "Rainforest Alliance" or "Sustainable Agriculture Network" or "UTZ Certified" or "UTZ" or "Global Partnership for Good Agricultural Practice" or "Global GAP" or "GlobalGAP" or "4C Association" or "Better Cotton Initiative" or "BCI" or "Cotton made in Africa" or Bonsucro or "Ethical Tea Partnership" or Trustea or "International Federation of Organic Agriculture Movements" or IFOAM or "soil association" or "IOAS" or "LEAF" or "Linking Environment and Farming" or "Union for Ethical BioTrade" or "UEBT" or "Roundtable on Sustainable Palm Oil" or "RSPO" "Fair Flowers Fair Plants" or "ProTerra" or "ISO 14001")) Show Less

S3

CC D18 or L15

S2

TI (((fair* OR ethic* OR alternative OR sustainab* OR responsib* OR specialty OR eco OR ecologic OR ecological OR organic) N3 (certifi* OR standard* OR label* OR seal* OR scheme* OR trad* OR market* OR "value chain*" OR commodit* OR product*))) OR AB (((fair* OR ethic* OR alternative OR sustainab* OR responsib* OR specialty OR eco OR ecologic OR ecological OR organic) N3 (certifi* OR standard* OR label* OR seal* OR scheme* OR trad* OR market* OR "value chain*" OR commodit* OR product*))) OR SU (((fair* OR ethic* OR alternative OR sustainab* OR responsib* OR specialty OR eco OR ecologic OR ecological OR organic) N3 (certifi* OR standard* OR label* OR seal* OR scheme* OR trad* OR market* OR "value chain*" OR commodit* OR product*))) Show Less

S1

TI (("certification" or "quality standards" or "quality label?ing" or "sustainability standards")) OR AB (("certification" or "quality standards" or "quality label?ing" or "sustainability standards")) OR SU (("certification" or "quality standards" or "quality label?ing" or "sustainability standards"))

Limiters - Published Date: 19900101-20150531

Example of Hand-Search Strings and Documentation

NB: This is a single Excel spreadsheet cut into three parts for legibility. The pictures align horizontally.

Website Searching Template	**2. S**
Name of Review: SR-1158	fill in
Name of website/source *select source from the dropdown list - you need to have filled in worksheet 1.Designing the search for this*	**Initia sear**
AgEcon #1	PFV
AgEcon #2	PFV
AgEcon #3	PFV
AgEcon #4	PFV
FairTrade Foundation #1	PFV
FairTrade Foundation #2	PFV
FairTrade Foundation #3	PFV
FairTrade Foundation #4	PFV
FairTrade Foundation #5	PFV
FairTrade Foundation #6	PFV
FairTrade Foundation #7	PFV
FairTrade Foundation #8	PFV

2. Searching

fill in all details of the search as appropriate

is	Initials of searcher	Date of first search	Date of last search (if different)	Pathway followed e.g. Browsed headings/searched site/database within website - *can use separate lines for the different types of searches. Please give suficient detail to allow for the replication of the search*	key app sea
	PFV	14/07/2015	15/07/2015	Search bar at top of home page (immediately visable) > Enter a random search term into search bar and press enter > this will take you to a more advanced search page > enter search terms as decribed in next column.	Sea
	PFV	14/07/2015	15/07/2015	Search bar at top of home page (immediately visable) > Enter a random search term into search bar and press enter > this will take you to a more advanced search page > enter search terms as decribed in next column.	Sea
	PFV	15/07/2015	17/07/2015	Search bar at top of home page (immediately visable) > Enter a random search term into search bar and press enter > this will take you to a more advanced search page > enter search terms as decribed in next column.	Sea
	PFV			Search bar at top of home page (immediately visable) > Enter a random search term into search bar and press enter > this will take you to a more advanced search page > enter search terms as decribed in next column.	Sea
	PFV	20/07/2015	20/07/2015	browse front page headings, select "**resources library**" > scroll down and browse categories available on page, select "**reseaching**" tab > scroll down and browse categories, select "**monitoring & impact**" tab.	n/a
	PFV	20/07/2015	20/07/2015	browse front page headings, select "**resources library**" > scroll down and browse categories available on page, select "**reseaching**" tab > scroll down and browse categories, select "**commodity information**" tab.	n/a
	PFV	20/07/2015	20/07/2015	browse front page headings, select "**resources library**" > scroll down and browse categories available on page, select "**reseaching**" tab > scroll down and browse categories, select "**policy briefings & papers**" tab.	n/a
	PFV	20/07/2015	23/07/2015	browse front page headings, select "**resources library**" > scroll down and browse categories available on page, select "**reseaching**" tab > scroll down and browse categories, select "**Policy Reports**" tab.	n/a
	PFV	20/07/2015	20/07/2015	browse front page headings, select "**what is fairtrade?**"> scroll down page and select "**facts and figures**" tab > scroll down page and browse documents	n/a
	PFV	20/07/2015	20/07/2015	browse front page headings, select "**what is fairtrade?**"> scroll down page and select "**The Impact of our Work**" tab > scroll down page and select "**Impact Research and Evaluation Studies" tab**	n/a
	PFV	21/07/2015	23/07/2015	browse front page headings, select "**what is fairtrade?**"> scroll down page and select "**Social and Annual Reports**" tab	n/a
	PFV	20/08/2015	20/08/2015	browse front page headings, select "**media centre**" > select "**press releases**" tab	sea

	keywords and fields searched - *can paste strategy here if appropriate. use separate lines for the different types of searches*	**Used defined keywords - "Y"**	**Number of documents found** - *use sepearate lines for each indidivual search*	**Additional notes**
	Search: All of AgEcon Search > **For:** ((keyword:certification) OR (keywo	N	257	
	Search: All of AgEcon Search > **For**:((Title:certification) OR (Title:quality	N	6	
	Search: All of AgEcon Search > **For:** (keyword:fair keyword:trade) OR (ke	N	754	
	Search: All of AgEcon Search > **For:** ((keyword:fair*) OR (keyword:ethic*	N	984	
	n/a - browse documents available on page	y	2	http://www.fairtrade.org.uk/en/resources-lib
	n/a - browse documents available on page	y	2	http://www.fairtrade.org.uk/en/resources-lib
	n/a - browse documents available on page	y	4	http://www.fairtrade.org.uk/en/resources-lib
	n/a - browse documents available on page	y	2	http://www.fairtrade.org.uk/en/resources-lib
	n/a - browse documents available on page	y	6	http://www.fairtrade.org.uk/en/what-is-fairtr
	n/a - browse documents available on page	y	6	http://www.fairtrade.org.uk/en/what-is-fairtr
	n/a - browse documents available on page	y	22	http://www.fairtrade.org.uk/en/what-is-fairtr
	search press releases available on page for relevant documents	n	115	http://www.fairtrade.org.uk/en/media-centre

Annex B: Included study descriptions for review questions 1 and 2

Descriptive summary of studies included for review question 1

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Anteneh et al 2014	Ethiopia	Coffee	Fairtrade; Organic, UTZ	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	700	1:FT&org; 2:FT&org&Utz	No	Average Price	High
Asfaw et al 2008	Kenya	Vegetables (French beans, green beans, peas)	EurepG.A.P./ GlobalGAP	Agricultural producers	Ex-post controlled observational study	2SLS/3SLS (Poisson, treatment effects); PSM	439	1:GlobalGAP	Yes	Illness; Net Income Cert Prod	High
Balineau 2012	Mali	Cotton	Fairtrade	Agricultural producers	Retrospective panel built from single data collection exploiting staggered intervention	Fixed effects panel model (GMM estimator)	198	1:FT (in four waves)	Yes	Quality	High
Barham & Weber 2012	Mexico; Peru	Coffee	Fairtrade; Organic; Rainforest Alliance	Agricultural producers	Panel build from two datasets	Fixed effects panel model	845(Mexico); 235 (Peru)	1:RA (Peru); 2: FT&org (Mexico)	No	Yield	Critical
Becchetti & Gianfreda 2008	Kenya	Fruit	Fairtrade; Organic	Agricultural producers	Ex-post controlled observational study	2 stage treatment model (ordered probit/probit and selection model)	478	1:FT&org; 2:Conversion to org (no FT); 3:Coop members only (no CS)	No	Illness Index	High

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Becchetti et al 2008	Chile	Honey	Fairtrade; Organic	Agricultural producers	Ex-post controlled observational study	GMM regression with LDVs (and a FE model)	234	1:FT	No	Productivity (income per h worked)	High
Becchetti et al 2011	Thailand	Rice	Fairtrade; Organic	Agricultural producers	Retrospective panel: ex-post controlled observational study with recall questions	3SLS	360	1:FT&org; 2:Org only	No	HH Income	High
Cepeda et al 2013	Ecuador	Cocoa	Rainforest Alliance; Organic; Fairtrade	Agricultural producers	Controlled before and after (CBA, quasi-experiment with baseline and endline data collection)	Difference-in-difference / double difference (DID	2010: 576 2012: 415	1:RA&org (later org only); 2:FT&org	Yes	None	Critical
Chiputwa & Qaim 2014	Uganda	Coffee	Fairtrade; Organic, UTZ	Agricultural producers	Ex-post controlled observational study	IV model, and simultaneous eqn. model with IV, estimated using maximum likelihood	419	1:FT&Utz; 2:Org&Utz	Yes	Calorie Consumption; P/C Cons. Exp.; Poverty Headcount	Moderate

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Colen et al 2014	Senegal	Green bean, mango	GlobalGAP	Wage workers	Controlled before and after (quasi-experiment with baseline and endline data collection)	Cross-sectional OLS; FE panel regression	163 (for OLS), 46 (for panel)	1:GlobalGAP	Yes	Daily Wages	Moderate
Bennett et al 2012	Cote d'Ivoire	Cocoa	Rainforest Alliance	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	PSM and DID	2009: 200; 2011: 252 (prior to matching)	1:RA	No	Yield; Revenue Per Ha; Net Income Per Ha; School Attendance	Moderate
Cramer et al 2014	Ethiopia	Coffee	Fairtrade	Wage workers	Ex-post controlled observational study	Robust OLS; Propensity score matching (PSM)	339	1:FT	Yes	Daily Wages	Moderate
Cramer et al 2014	Ethiopia	Flowers	Fairtrade	Wage workers	Ex-post controlled observational study	Robust OLS; Propensity score matching (PSM)	215	1:FT	Yes	Daily Wages	Moderate
Cramer et al 2014	Uganda	Coffee	Fairtrade	Wage workers	Ex-post controlled observational study	Robust OLS; Propensity score matching (PSM)	237	1:FT	Yes	Daily Wages	Moderate

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Cramer et al 2014	Uganda	Tea	Fairtrade	Wage workers	Ex-post controlled observational study	Robust OLS; Propensity score matching (PSM)	199	1:FT	Yes	Daily Wages	Moderate
Dragusanu 2014	Costa Rica	Coffee	Fairtrade	Agricultural producers	Ex-post controlled observational study	Linking of mill-level data with individual HH survey data to estimate FT impact	Unclear, but very large	1:FT	Yes	Wage Income	High
Ehlert et al. 2014	Kenya	Vegetables	GlobalGAP	Wage workers	Ex-post controlled observational study	OLS and MIMIC models	316	1:GlobalGAP small farms; 2:GlobalGAP large farms	Yes	Hourly Wages	High
Fort & Ruben 2009	Peru	Banana	Fairtrade	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	200	1:FT&org	No	Net Income Cert. Prod; Measure Of Wealth	Moderate
García et al 2014	Colombia	Coffee	UTZ	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	PSM and DID	857, much smaller for PSM though	1:Utz	No	None	Moderate
Jena et al 2012	Ethiopia	Coffee	Fairtrade; Organic	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	249	1:FT&org	Yes	P/c Gross Income; Total HH Income; Yield	High

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Kamau et al 2010	Kenya	Coffee	UTZ	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	262, after matching	1:Utz	No	None	High
Kuit et al 2016	Uganda	Coffee	4C	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	250	1:4C	No	None	High
Kuit et al 2016	Vietnam	Coffee	4C	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	Propensity score matching (PSM)	2006-2008: 152; 2013-2015: 249	1:4C	No	None	High
Minten et al 2015	Ethiopia	Coffee	Various	Agricultural producers	Ex-post controlled observational study	Robust OLS; Propensity score matching (PSM)	1600	1:Various (assignment to CS not possible)	Yes	Child Labour; Schooling	Moderate
Mueller & Theuvsen 2015	Guatemala	Peas	GlobalGAP	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	276	1:GlobalGAP	Yes	Total HH Income; Rev From Cert. Prod.	Moderate
Nelson et al 2013	Ghana	Cocoa	Fairtrade	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	PSM and DID	2010: 743, 2012: 697	1:FT	Yes	None	Critical

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Parvathi & Waibel 2016	India	Black pepper	Fairtrade; UTZ; Organic	Agricultural producers	Ex-post controlled observational study	Multinomial endogenous switching model	300	1:FT&org	Yes	Total Income P/C; Assets Per Capita	High
Riisgaard et al 2009	Uganda	Coffee	Fairtrade	Agricultural producers	Ex-post controlled observational study	2-stage treatment model with both OLS and FIML	149	1:FT&org	No	Net Revenue Cert Prod	Moderate
Riisgaard et al 2009	Uganda	Coffee	UTZ	Agricultural producers	Ex-post controlled observational study	2-stage treatment model with both OLS and FIML	147	1:Utz&org	No	Net Revenue Cert Prod	Moderate
Roy & Thorat 2008	India	Grapes	EurepGAP/ GlobalGAP	Agricultural producers	Ex-post controlled observational study	2SLS (IV)	183	1:FT	No	None	High
Ruben & Fort 2012	Peru	Coffee	Fairtrade; Organic	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	360	1:FT&org; 2:FT only	No	Gross Coffee Income; Coffee Profits; Price; Yield; Total HH Gross Income	Moderate
Ruben & Zúñiga-Arias 2011	Nicaragua	Coffee	Fairtrade; Organic	Agricultural producers	Ex-post controlled observational study	Propensity score matching (PSM)	315	1:FT; 2:FT&org	Yes	None	Unclear

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Ruben et al 2014	Kenya	Coffee	Fairtrade; UTZ	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	PSM and DID	2009: 600; 2013: 493	1:FT; 2:Utz	No	None	High
Schuster & Maertens 2014	Peru	Horticulture (asparagus, grapes, avocado, artichoke)	Various (29 different private standards)	Wage workers	Controlled before and after (CBA, quasi-experiment with baseline and endline data collection)	Instrumental variable	2013: 592; 2014: 499	1:Various (assignment to CS not possible)	Yes	Hourly Wage; Length Of Employment	Low
Stathers & Gathuthi 2013	Kenya	Tea	Fairtrade, Rainforest Alliance	Agricultural producers	Controlled before and after (CBA, quasi-experiment with baseline and endline data collection)	Difference-in-difference	2010: 897; 2012: 700	At endline: 1:FT&RA; 2:FT; 3: FT but no coop; 4:RA (large farms)	Yes	None	Critical
Subervie & Vagneron 2013	Madagascar	Lychee	GlobalGAP	Agricultural producers	Ex-post controlled observational study	PSM and DID	505	1:GlobalGAP current; 2:GlobalGAP former	Yes	Max Price Received	High
van Rijn 2016	Ghana	Banana	Fairtrade	Wage workers	Ex-post controlled observational study	Propensity score matching (PSM)	326	1:FT	No	None	Critical

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
van Rijn 2016	Dominican Republic	Banana	Fairtrade	Wage workers	Ex-post controlled observational study	Propensity score matching (PSM)	258	1:FT	No	None	High
van Rijn 2016	Colombia	Banana	Fairtrade	Wage workers	Ex-post controlled observational study	Propensity score matching (PSM)	431	1:FT	No	None	High
Van Rijsbergen et al 2016	Kenya	Coffee	Fairtrade; UTZ	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	PSM and DID	218	1:FT; 2:FT&Utz	No	Gross Coffee Income; Yields	Low
Waarts et al 2012	Kenya	Tea	Rainforest Alliance	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	Difference-in-difference (DID)	331	1:RA&Farmer field schools; 2:RA; 3:Farmer field schools only	Yes	Net Income	High

Study	Location	Com-modity	Certification scheme(s)	Popula-tion	Study design	Method of analysis	Sample size	Treatment arms	Indepen-dently financed?	ES extracted	Overall risk of bias
Waarts et al 2016	Ghana	Cocoa	UTZ	Agricultural producers	Controlled before and after (quasi-experiment with baseline and endline data collection)	PSM and DID	352	1:Utz	No	Productivity Per Ha; Profit Per Ha; Net Income Cert Prod.; Total HH Income	High
Weber 2011	Mexico	Coffee	Fairtrade; Organic	Agricultural producers	Ex-post controlled observational study	Treatment effects model estimated with maximum likelihood	845	1:FT&org	1	Coffee Price	Moderate

Descriptive summary of included studies for review question 2

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Abarca-Orozco 2015	n/a	Fairtrade	Material related to: training and new practices; financial premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation; markets; production cost for certified goods;	Non-ethnographic. Semi-structured interviews; participant observation.	• A combination of bad management, failure to receive a FT price and loans can lead to a cooperative nightmare • Good relationships with a supportive buyer can make a huge difference in the effectiveness of FT • Barriers to adopting FT-organic are small farm and HH size and low education, and farm management requirements
Aidenvironment 2016	n/a	UTZ	Material related to: training and new practices	None	Material related to: participants' reception/motivation; markets; production cost for certified goods	Non-ethnographic. Mixed methods.	• Uptake of new practices from training is mediated by market demand for these practices • Strong public extension services and support can be a facilitator • FT benefits were found in the wider community beyond certified farmers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Amekawa 2001	n/a	Q-GAP (GlobalGAP)	None	None	Material related to: effective/non-effective adoption; markets	Non-ethnographic. Structured interviews	• Participants' lack of understanding and awareness of certification may be a barrier to compliance and uptake • Q-GAP's criteria on pesticides resulted in blemished fruit and reduced exportability, inhibiting uptake of standard • Producers saw a lack of demand for certified fruit and minimal commercial benefits
Arce 2009	n/a	Fairtrade, organic	Material related to: training and new practices	Material related to: distribution of benefits & investments	Material related to: production cost for certified products	Ethnographic.	• Costs of certified production can be a barrier, particularly those households with older members or those who have lost household labour due to migration • Reluctance of farmers to adopt new practices such as cutting down trees • Support from church/NGOs helped formation of coops and adoption of certification

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Asfaw et al 2009	n/a	GlobalGAP	Material related to: implementation costs of certification programmes	None	Material related to: cooperative management/performance; effective/non-effective adoption	Non-ethnographic. Questionnaires, interviews, group discussions.	• Investment in infrastructure and equipment made up most of the costs are incurred during implementation • Wealthier and more educated householders were more likely to adopt certification
Babin 2012	n/a	Fairtrade	None	None	Material related to: cooperative management/performance; participants' reception/motivation	Ethnographic.	• Poor cooperative performance hinders success, especially high cooperative administrative costs, cooperative debt in relation to lack of financial infrastructure, and abusive credit interest rates
Bacon 2005	n/a	Fairtrade, organic	Material related to: training and new practices	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; markets;	Non-ethnographic. Surveys.	• Reports on limited women participation; aid received by the third level coop; importance of delay payments; non-certified farmers have lower levels of education; organic requires more work
Bagama et al 2014	n/a	UTZ	Material related to: training and new practices	None	Material related to: cooperative management/performance; participants' reception/motivation	Non-ethnographic. Survey, focus group discussions, interviews.	• Premium prices are a key motivation for joining, while high yields are a key motivation for staying Also highlights increased yields as a result of certification good practices

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Bakker 2014	n/a	Rainforest Alliance	Material related to: participant's targeting and self-selection	Material related to: uneven adoption of standard	Material related to: participants' reception/motivation; markets, production cost for certified goods	Non-ethnographic. Interviews, observation, some secondary documentation research	• The factory selects which farmers will be certified, which meant that large farmers already closest to meeting requirements were most likely to be certified and smallholders were disadvantaged • Clearer communication between farmer and factory around harvest, lead to better quality at time of sale
Balineau 2011	n/a	Fairtrade	Material related to: participant's targeting and self-selection	None	Material related to: participants' reception/motivation	Non-ethnographic. Survey.	• Literacy is an important barrier for administrative posts in the FT coop. • This results in the main paperwork load falling on the few literate/French speaking producers within the PO.
Barham and Weber 2012 (Mexico)	n/a	Rainforest Alliance	Material related to: training and new practices	None	None	Non-ethnographic. Survey data, cooperative records.	• Yield, rather than price premiums (which were marginal), is a determinant of higher net household returns • Practices that contributed to higher yields included systematic pruning and appropriate fertilising

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
arham and Weber 2012 (Peru)	n/a	Rainforest Alliance	Material related to: training and new practices	None	None	Non-ethnographic. Mixed methods: Survey data, cooperative records.	• Yield, rather than price premiums (which were marginal), is a determinant of higher net household returns • Practices that contributed to higher yields included systematic pruning and appropriate fertilising
Beall 2012	n/a	RSPO	Material related to: training and new practices; participant's targeting and self-selection; implementation costs of certification programmes	None	Material related to: participants' reception/motivation; social infrastructure; demand for certified products	Non-ethnographic."Country level assessments" and semi-structured interviews.	• There were not many incentives for certification at the moment, due to high costs, lack of infrastructure to assist with certification training and process, and overcapacity in mills. • Current project's success is due to outside aid and support, but no group seems prepared to step in with help on costs and training, and without that incentives are unlikely to be enough to convince farmers to join

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Bergeron 2010	n/a	Fairtrade	None	Material related to: uneven adoption of standard	Material related to: participants' reception/motivation	Non-ethnographic. Semi-structured interviews.	• Female participation is related to opportunities for income diversification or family ties and not to the Fairtrade certification • The existence of purely female groups/coops facilitates women's participation. • Barriers include education and lack of knowledge and age of children, distance/lack of transport from the coop, late hours of meetings
Besky 2014	n/a	Fairtrade	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation; governmental legislation, incentives, and/or regulation	Ethnographic.	• Highlights conflict between Fairtrade's categorisation of plantation workers as hired "farm workers" and its "farm-worker standards" and the reality of the historic plantation system, in which home is work for labourers, and "facilities" necessarily comprise part of the compensation • Similarly, the mismatch between FT's market-driven methods and the plantation institution

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Beuchelt 2009	n/a	Fairtrade, organic	Material related to: financial premium use	None	Material related to: cooperative management/performance; markets; production cost for certified goods	Non-ethnographic. Semi-structured interviews.	• Lack of transparency regarding final prices enables easy misuse of funds • For certified cooperatives high conventional coffee prices are a threat because farmers increase sales to conventional market channels
Bonanno and Cavalcanit 2012	n/a	GlobalGAP	Material related to: training and new practices	None	Material related to: governmental legislation, incentives, and/or regulation	Non-ethnographic. Case study methodology-observations, interviews, documents review	• Highlights the shifts in labour conditions, remuneration, and hours/seasons • The real driver is not GLOBALGAP certification, but the demands producers must comply with to supply global supermarkets, and the labour effects that come from supplying this market
Brown 2012	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments	Material related to: markets	Non-ethnographic. Interviews.	• Reports on the potential negative spill over effects on trade unions, as premiums may make workers less eager to fight for their rights through unions. Also on possible conflicts regarding premiums management and tension between certified and non-certified producers.

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Carimentrand and Ballet2010	n/a	Fairtrade, organic	None	Material related to: distribution of benefits & investments	None	Non-ethnographic. Interviews.	• In an increasingly unequal economic context of quinoa production in Bolivia, FT certification benefits the larger producers more than the smaller ones.
Cepeda et al 2013	n/a	Fairtrade, organic	Material related to: training and new practices; certification related services; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; participants' reception/motivation; social infrastructure; markets; production cost for certified goods	Non-ethnographic. Survey, questionnaires, interviews, focus groups.	• The FT audit process enhances transparency for the certified organisations as the findings of the audits are made available to all member • Organic farmers had dropped Rainforest Alliance certification because they felt organic gave them a premium price and more assured market
CESU 2012 (Ecuador)	n/a	Fairtrade	Material related to: training and new practices; certification related services; social premium use	Material related to: distribution of benefits & investments	None	Non-ethnographic. Survey, observation, interviews, focus groups.	• Women's committees existed in both coops and supported female family members of farmers achieve greater independence through income • Female farmers do not have the time to participate in the committee meeting, so FT helped female family members more than female farmers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
CESU 2012 (Kenya)	n/a	Fairtrade	Material related to: training and new practices; certification related services; social premium use	Material related to: distribution of benefits & investments	Material related to: social infrastructure	Non-ethnographic. Survey, observation, interviews, focus groups.	• Workers needed to pay half the price for popular trainings, limiting their accessibility • Credit scheme financed through premium was viewed by workers as an important advantage
CESU 2012 (Ghana)	n/a	Fairtrade	Material related to: training and new practices; certification related services; financial premium use; social premium use	Material related to: distribution of benefits & investments	None	Non-ethnographic. Survey, observation, interviews, focus groups.	• With trainings, literacy can be a barrier to teaching complex details and time/financial constraints can be a barrier to attendance • Low level of education makes writing proposals for spending premium difficult for the farmers; farmers also had low understanding of process
CESU 2012 (Peru; Coffee)	n/a	Fairtrade	Material related to: training and new practices; certification related services; financial premium use; social premium use	Material related to: elite capture	Material related to: cooperative management/performance; social infrastructure	Non-ethnographic. Survey, observation, interviews, focus groups.	• Premium was used in times of high coffee prices to help coops pay higher prices to their farmers • Peru deducts part of the premium as tax

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
CESU 2012 (Peru; Banana)	n/a	Fairtrade	Material related to: training and new practices; social premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation; social infrastructure	Non-ethnographic. Survey, observation, interviews, focus groups.	• Women's committees gave women opportunities to be trained on practical topics like handicraft or cooking • Premium was used in times of high coffee prices to help coops pay higher prices to their farmers • Peru deducts part of the premium as tax
CESU 2012 (India)	n/a	Fairtrade	Material related to: training and new practices; social premium use	Material related to: distribution of benefits & investments	None	Non-ethnographic. Survey, observation, interviews, focus groups.	• Cultural reasons meant there were very few female farmers, so women had no voice in decision-making process on premium usage
Chiputwa et al 2015	n/a	Fairtrade, organic, UTZ	Material related to: participant's targeting and self-selection	None	Material related to: cooperative management/performance; markets	Non-ethnographic. Interviews with structured questionnaire.	• Processing by coop before selling coffee may add value and increase prices • FT coops own the certification papers in FT, increasing their autonomy in marketing and who to sell to; with UTZ and Organic the exporters hold the certificates, meaning farmers must sell to specific buyers and have less marketing freedom

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Cofre et al 2012	n/a	GlobalGAP	Material related to: participant's targeting and self-selection; monitoring and auditing; implementation costs of certification programmes	Material related to: uneven adoption of standard	None	Non-ethnographic. Questionnaire-based sample survey.	• Key data show that GAP certification costs are very high, but variable according to farm size and other characteristics • The certification is especially suitable to well-established agribusiness, which can bear the high costs of introduction of certification and its maintenance • Perception of benefits are strong but this is also because this certification is seen as a necessary condition to access the best markets
Cramer et al 2014a (Ethiopia; Coffee)	n/a	Fairtrade	None	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Non-ethnographic. Mixed methods: surveys, questionnaires, life's work interviews.	• FT did not result in better working conditions, longer job duration, or higher wages for wage workers • The heterogeneity of "smallholders" ignores the dependence of the poor on wage labour incomes and means that interventions designed to increase all participating famers' income will not successfully reduce poverty • Large smallholders may help reduce poverty through their use of wage workers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Cramer et al 2014a (Ethiopia; Flowers)	n/a	Fairtrade	None	Material related to: distribution of benefits & investments	None	Non-ethnographic. Mixed methods: surveys, questionnaires, life's work interviews.	• FT did not result in better working conditions, longer job duration, or higher wages for wage workers • The heterogeneity of "smallholders" ignores the dependence of the poor on wage labour incomes and means that interventions designed to increase all participating famers' income will not successfully reduce poverty • Large smallholders may help reduce poverty through their use of wage workers
Cramer et al 2014a (Uganda; Coffee)	n/a	Fairtrade	None	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Non-ethnographic. Mixed methods: surveys, questionnaires, life's work interviews.	• FT did not result in better working conditions, longer job duration, or higher wages for wage workers • The heterogeneity of "smallholders" ignores the dependence of the poor on wage labour incomes and means that interventions designed to increase all participating famers' income will not successfully reduce poverty • Large smallholders may help reduce poverty through their use of wage workers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Cramer et al 2014a (Uganda; Tea)	n/a	Fairtrade	None	Material related to: distribution of benefits & investments	None	Non-ethnographic. Mixed methods: surveys, questionnaires, life's work interviews.	• FT did not result in better working conditions, longer job duration, or higher wages for wage workers • The heterogeneity of "smallholders" ignores the dependence of the poor on wage labour incomes and means that interventions designed to increase all participating famers' income will not successfully reduce poverty • Large smallholders may help reduce poverty through their use of wage workers
Cramer et al 2014b (Uganda)	n/a	Fairtrade	Material related to: certification related services; social premium use	Material related to: distribution of benefits & investments; elite capture	Material related to: effective/non-effective adoption	Non-ethnographic. Mixed methods: market data analysis and interviews.	• Interventions aimed at cooperatives will not be successful in poverty alleviation - rather interventions must be focused on promoting a more rapid rate of growth in wage labour in export crop production • Elite capture is perpetuated through the fact that subsidies and size/volume of sales are directly correlated to positions of power within coop

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Cramer et al 2014b (Ethiopia)	n/a	Fairtrade	Material related to: certification related services; social premium use	Material related to: distribution of benefits & investments	Material related to: effective/non-effective adoption	Non-ethnographic. Mixed methods: market data analysis and interviews.	• Interventions aimed at cooperatives will not be successful in poverty alleviation - rather interventions must be focused on promoting a more rapid rate of growth in wage labour in export crop production • Elite capture is perpetuated through the fact that subsidies and size/volume of sales are directly correlated to positions of power within coop
Dannenberg and Nduru 2013	n/a	GlobalGAP	Material related to: participant's targeting and self-selection; implementation costs of certification programmes	None	Material related to: effective/non-effective adoption	Non-ethnographic. Mixed methods: quantitative survey and interviews.	• Found serious barriers to GlobalGAP adoption - cost and complexity - but that many horticultural producers were able to access the market through other routes
Dolan 2010	n/a	Fairtrade	Material related to: monitoring and auditing; implementation costs of certification programmes; use of financial premiums; use of social premiums	Material related to: elite capture	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation; markets	Non-ethnographic.	• The uncoupling of social premium projects from their institutional context creates problems such as bad decisions in the use of social premium and substantial potential for elite capture, perception of little benefits by many farmers, and ineffective auditing processes

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Donovan and Poole 2014a	Donovan 2014b	Fairtrade, organic	Material related to: training and new practices; certification related services	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; markets	Non-ethnographic. Survey, interviews.	• Payment delays and need to commute to coop offices make participation in the FT coop costly • Access to credit with extended repayment periods played a critical role in expanding and improving natural capital. • Most households acquired new skills that improved coffee quality, but few households had acquired the more complex skills for improved plantation management
Donovan and Poole 2014b	Donovan 2014a	Fairtrade, organic	Material related to: training and new practices; certification related services	None	Material related to: cooperative management/performance; markets	Non-ethnographic. Key informant interviews, household surveys, secondary information.	• As above, and: credit can be positive but also negative: producers do sell to the coop if they cannot afford tor repay the debt

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Dowdall 2012	n/a	Fairtrade, organic	Material related to: training and new practices; certification related services; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: cooperative management/performance; markets; market volatility; production cost for certified goods	Ethnographic.	• Certified production is costly and not all producers can afford it • FT loses advantage when conventional prices rise and overpass FT prices - premiums were used to supplement FT prices in order to compete with conventional prices • Producers without enough education/resources to obtain other sources of income struggle to participate in certified markets
Dragusanu and Nunn 2014	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes	Material related to: distribution of benefits & investments; uneven adoption of standard	None	Non-ethnographic. Secondary archival data.	• FT was only associated with increased incomes for a small group of skilled coffee growers and farm owners • FT certification was associated with increased school attendance in the region • FT certification was associated with lower school attendance among some children of coffee workers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Ellery 2010	n/a	Fairtrade, organic	Material related to: financial premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance	Non-ethnographic. Semi-structured interviews.	• Women-only sub-coop was to be formed in order to give participation space to women • Women-only FT coops help to visibilise the work of women in coffee production
Fairtrade 2013 (Tea)	n/a	Fairtrade	Material related to: implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; social infrastructure; markets; production cost for certified goods	Non-ethnographic. Participatory livelihoods approach, focus group discussions, SSI.	• FT does not seem to significantly alter price conditions and marketing effects but is largely valued for the contribution of social premium projects and the good leadership in organisations

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Fairtrade 2013 (Groundnuts)	n/a	Fairtrade	Material related to: training and new practices; certification related services; implementation costs of certification programmes	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; social infrastructure; markets; production cost for certified goods	Non-ethnographic. Participatory livelihoods approach, focus group discussions, SSI.	• Disappointment with the size of the premium and weaknesses in the purchasing system, but praise for extension work • Since extension services and the existence of the group predates the certification it seems that the main benefits cannot be attributed to FT per se. If anything a strengthening of existing capacities is suggested • This is another story of context in which a key actor and its limitations severely constrain the potential benefits of FT certification

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Fairtrade 2013 (Sugar)	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation; markets; production cost for certified goods	Non-ethnographic. Participatory livelihoods approach, focus group discussions, SSI.	• The case of this sugar coop shows how FT benefits in terms of premium income and social premium are particularly appreciated in difficult times • Debts were incurred due to the Phase II expansion of the scheme, and is expected to be settled in 2016 when economies of scale are reaped • Lack of control comes from the fact that farmers are de facto tenants of the Shire Valley Cane Grower Trust Board with which relations are tense but FT is unable to mediate thus far
Fairtrade 2015 (Bananas)	n/a	Fairtrade	Material related to: participant's targeting and self-selection	Material related to: distribution of benefits & investments; elite capture	Material related to: cooperative management/performance; participants' reception/motivation	Non-ethnographic. Focus group discussions, SSI, LH.	• There are important barriers to women's participation as well as risks for the enhancement of their position given that barriers are deep-rooted and require context-specific steps to tackle • There is a limit to how much can be done through FT to reduce such barriers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Fairtrade 2015 (Cotton)	n/a	Fairtrade	Material related to: implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments; elite capture	Material related to: cooperative management/performance; participants' reception/motivation	Non-ethnographic. Focus group discussions, SSI, LH.	• There are important barriers to women's participation as well as risks for the enhancement of their position given that barriers are deep-rooted and require context-specific steps to tackle • There is a limit to how much can be done through FT to reduce such barriers
Fairtrade 2015 (Tea)	n/a	Fairtrade	Material related to: implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments; elite capture	Material related to: cooperative management/performance	Non-ethnographic. Focus group discussions, SSI, LH.	• There are important barriers to women's participation as well as risks for the enhancement of their position given that barriers are deep-rooted and require context-specific steps to tackle • There is a limit to how much can be done through FT to reduce such barriers
Fayet and Vermeulen 2014	n/a	Fairtrade, organic, Better Cotton Initiative, Shop for Change	Material related to: implementation costs of certification programmes	None	Material related to: markets	Non-ethnographic. Interviews.	• Company based certification projects provide better market access than NGO based initiatives

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Franccesconi and Ruben 2014	n/a	Fairtrade	Material related to: Participant's targeting and self-selection; certification related services; monitoring and auditing; implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation; markets	Non-ethnographic. Quantitative survey.	• Over time, increased side-selling and over-certification tend to reduce the positive effects of FT premia and even neutralise them • For FT to maximise outcomes, it should be targeted to particularly well-designed coops
Fraser et al 2014	n/a	Fairtrade	Material related to: social premium use	Material related to: elite capture; uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation	Ethnographic.	• Local politics/relations can affect access to Fairtrade. In the case studied, the FT cooperative has been hijacked by the local elite and excluded the landless group, affecting farmers' trust in Fairtrade and coops and affecting producers' returns • It also highlights how "Fair trade" values and possible access to fair trade markets motivated producers to repay the debt and consolidate their organisation

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Garcia et al 2014	n/a	UTZ	Material related to: training and new practices; implementation costs of certification programmes	None	Material related to: cooperative management/performance; participants' reception/motivation; demand for certification services	Non-ethnographic.	• Highlights positive perceptions of the programme but also problems with implementation, including failure to continue training • It also found higher income and lower cost per kilo, mainly due to higher yields
Getz and Schreck 2006	Schreck 2002	Fairtrade	Material related to: training and new practices	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation; markets	Non-ethnographic. Questionnaire.	• Certification was driven by a process implemented by the buying company, and producers totally lacked understanding and knowledge about how FT works • There was uneven access to the FT market within the coop based on quality criteria
Gómez-Cardona 2012	n/a	Fairtrade, organic	Material related to: training and new practices; monitoring and auditing; implementation costs of certification programmes	None	Material related to: markets	Ethnographic.	• Certification led to formalising productive activities and also to an increase in bureaucracy • Barriers included certification paperwork, which was a considerable workload for producers, and increasingly strict quality requirements

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Hanson et al 2012	n/a	Fairtrade	None	Material related to: uneven adoption of standard	Material related to: cooperative management/performance	Non-ethnographic. Participant observation, semi-structured interviews, gender-disaggregated focus groups.	• Women-only coops emerge as a reaction by women to being underserved by existing institutions • FT results in a double burden of work for women as it does not address domestic work burden distribution. • Gender equity goes beyond women's participation – which has been the de facto indicator of equity in most existing research
Heller 2010	n/a	Fairtrade, organic, UTZ, C.A.F.E Practices	Material related to: certification related services; monitoring and auditing; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation; governmental legislation, incentives, and/or regulation; markets; demand for certified products	Ethnographic.	• Despite standards requirements producers struggle to pay minimum wage, and wage workers get the local "jornal" • Producers are being "warned and prepared by the central coop " before audits • Registration bureaucratic requirements are difficult if not impossible to apply correctly
Herman 2010	n/a	Fairtrade	Material related to: social premium use	Material related to: distribution of benefits & investments	Material related to: social infrastructure	Non-ethnographic. Interviews.	• The distributional dynamics created between Fairtrade and non-FT farms has the potential for "raising the bar" and spurring other farms to certify

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Jaffee 2006	n/a	Fairtrade, organic	Material related to: training and new practices; certification related services; implementation costs of certification programmes	None	Material related to: cooperative management/performance; markets	Ethnographic.	• Gov't support programmes organised through the FT coop can be the strongest motivation to join the FT coop • Organic certification requirements have increased the workload of both men and women and the use of hired labour, which can be a barrier to participation • The FT increased labour requirement created wage employment
Jari et al 2013	n/a	Fairtrade	Material related to: participant's targeting and self-selection; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; markets	Non-ethnographic. Interviews.	• Highlights advantages (investments in local community development) and disadvantages (competition with small-scale farmers) to including commercial farms in FT • Makes an argument for certifying commercial farms – a lack of certification would hurt farm workers (who benefit from premium-funded community development projects) more than it would hurt the farm owners

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Jena et al 2012	n/a	Fairtrade, organic	Material related to: training and new practices; certification related services; social premium use	None	Material related to: cooperative management/performance; participants' reception/motivation	Non-ethnographic. Interviews, survey, focus group discussions.	• There is high heterogeneity between certified coops, and a producer's specific coop can make a big difference on things like trainings and access to credit
Kariuki 2014	n/a	GlobalGAP	None	None	Material related to: effective/non-effective adoption	Non-ethnographic. Interviews and questionnaires.	• Factors correlated with successful certification include size of land under cultivation, number of extension visits, and farm assets
Köhne 2014	n/a	RSPO	Material related to: monitoring and auditing	None	Material related to: governmental legislation, incentives, and/or regulation	Non-ethnographic. Interviews, participatory observation, archival research.	• RSPO as a multi-stakeholder initiative may be used by local landowners (who are uncertified) in disputes with large companies/plantations (who are RSPO certified or members). But power dynamics, heavily weighted towards the large companies, play a large role in the success of these disputes and in the success of deploying RSPO to adjudicate.

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Laroche et al 2012	n/a	Fairtrade	Material related to: certification related services; monitoring and auditing; implementation costs of certification programmes; financial premium use; social premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance	Non-ethnographic. Focus groups.	• Organic made a real difference in price, but with conventional FT and non-certified there was no significant difference; there was also no marked income difference for Fairtrade farmers compared to the rest of the population • Certifying in FT and organic makes production more technical, requiring more labour
Larsen et al 2014	n/a	RSPO	Material related to: monitoring and auditing	None	None	Non-ethnographic. Qualitative case-study methodology inspired by the principles of participatory action research.	• Highlights the difficulties in implementing and enforcing RSPO criteria on water rights and quality, resulting largely from lack of accountability (both in wider governmental authorities and within RSPO) and regulatory framework

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Loconto and Simbua 2010	n/a	Fairtrade, organic	Material related to: participant's targeting and self-selection; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; markets; demand for certified products	Non-ethnographic. SSIs, focus group discussions.	• Per cent of price going to producers may be lower in FT channels due to organisation of sales and collection in different chains and the weak distribution of benefits within FT • FT does not affect the price dynamics of conventional trade relations; pre-existing trade relations and their governance shape the outcomes of certification • Main problem is that FT benefits can also be obtained through conventional trade relationships so it is hard to disentangle the specific value added

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Luetchford 2008	n/a	Fairtrade	None	None	Material related to: participants' reception/motivation; markets	Ethnographic.	• Farmers expressed ambivalence towards the coop because of the capitalistic aspect of the coop - making profits, employing people for administrative and office work (not seen as work by farmers), acting as and dealing with intermediaries • The scale of the alternative market was limited, reducing efficacy of FT in terms of remuneration
Lyall 2014	n/a	Fairtrade	Material related to: training and new practice; implementation costs of certification programmes; social premium use	None	None	Non-ethnographic. Focus groups.	• Certification benefits such as minimum wages increases came with increased quotas and pressure on workers, as well as reductions in personnel • FT Premium is used as a mechanism to increase pressure on workers - risk of construing FT standards as "benefits" or "favours" that workers must repay one way or another

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Lyon 2005	Lyon 2006, Lyon 2007, Lyon et al 2010	Fairtrade, organic	Material related to: training and new practices; monitoring and auditing	None	Material related to: demand for certified products	Ethnographic.	• Irrational and excessive requirements included completely removing a popular tree because it was non-native • Lack of transparancy in auditing reports and bad communication of monitoring and auditing findings with producers
Lyon 2006	Lyon 2005, Lyon 2007, Lyon et al 2010	Fairtrade, organic	Material related to: training and new practices	Material related to: uneven adoption of standard	None	Ethnographic.	• Producers who were unable to read or write could not comply with standards, as they were unable to "maintain a daily written record of their agricultural activities"
Lyon 2007	Lyon 2005, Lyon 2006, Lyon et al 2010	Fairtrade, organic	Material related to: certification related services	None	None	Ethnographic.	• Having a commercial relationship, potentially due to FT, provided low interest loans that served to repay coop debt and provide credit with better terms to members

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Lyon et al 2010	Lyon 2005, Lyon 2006, Lyon 2007	Fairtrade, organic	None	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: participants' reception/motivation; production cost for certified goods	Non-ethnographic.	• FT-organic programs have increased women's participation but it is unclear whether this is to the benefit of women or just creating extra workload deriving from the more labour demanding FT-organic standards. • FT-organic programs increase the possibility that a woman is registered as a farm operator, which appeared to have positive effects in female participation/visibilisation and can be benefit in terms of recognising land property • Women maybe excluded from organisational posts due to language barriers or even mobility

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Makita 2011	n/a	Fairtrade	None	Material related to: uneven adoption of standard	Material related to: markets	Non-ethnographic. Semi-structured interviews.	• Small/marginal coffee farmers preferred diversification and crops with year-round income over FT coffee and its price volatility, labour costs and once yearly income • Organic certification was more attractive for subsistence farmers as they were already organic by default and certification did not mean any major changes in their farming practices. • FT was more suitable for wealthier part-time farmers who had already diversified income sources and were looking for intensifying their profits from agriculture
Makita 2012	n/a	Fairtrade	Material related to: social premium use	None	Material related to: cooperative management/performance	Non-ethnographic. Semi-structured interviews.	• Staff appointed by management, not workers, participate in the Joint Body • Lack of knowledge on source of FT premium projects can increase patron compliance as workers feel obliged to the patron for his generosity

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Melkeraaen 2009	n/a	Fairtrade	Material related to: implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments; elite capture	Material related to: cooperative management/performance	Non-ethnographic. Interviews.	• Language can be a barrier – FLO information is only available in English but most workers only speak Afrikaans • Company owners define the FLO-CERT regulation that workers must own 25% of the company differently – some as land ownership, others as brand ownership, which is seen as fulfilling the letter but not spirit of the law
Mendez 2002	n/a	Fairtrade, organic	Material related to: certification related services; financial premium use; social premium use	Material related to: elite capture; uneven adoption of standard	Material related to: cooperative management/performance; markets	Non-ethnographic. Focus groups.	• FT seems to enforce transparency and communication in coops • Inter-coop cooperation can be a facilitator, while second level coop fees can be a barrier for new small coops
Milford 2004	n/a	Fairtrade	Material related to: training and new practices; participant's targeting and self-selection; certification related services; financial premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation; markets; production cost for certified goods	Non-ethnographic. Interviews.	• Mexican coops are dependent on FT premium to operate • Organic production costs are a barrier to joining FT coops • Coop corruption and mismanagement a barrier to participation; effective coops indirectly (or directly) exclude producers by applying more strict entrance criteria

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Milford 2014	n/a	Fairtrade, organic	Material related to: participant's targeting and self-selection; certification related services	None	Material related to: cooperative management/performance; participants' reception/motivation; social infrastructure; markets, market volatility	Non-ethnographic. Interviews.	• Farmers with smaller farms, and smaller family labour force are less likely to join a FT cooperative due to organic production and coop-related costs • Producers who cannot afford payment delays or need large credits end up selling to the coyote. • Other barriers included politics, coop corruption and fines for not attending meetings
Minten et al 2015	n/a	Fairtrade, Organic, Various VSS	Material related to: participant's targeting and self-selection; financial premium use	None	Material related to: markets; demand for certified products	Non-ethnographic. Surveys and secondary databases.	• Frequent oversupply of certified products means that a lot of produce was sold on conventional markets • Producers received only 1/3 the premium that exporters of certified coffee received, in part because primary coops may keep some for overhead, or decide how to use the extra budget they receive from the premium

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Moberg 2008	Moberg 2005	Fairtrade	Material related to: participant's targeting and self-selection; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation; demand for certified products; production cost for certified goods	Ethnographic.	• The high costs of EurepGAP certification squeezed profit with no alternative (need to comply to export), while FT ban on herbicide increased wage costs for certified farmers • Lack of demand for FT bananas meant only a small percentage were actually sold as FT and received higher prices • Supermarket demands on provenance influenced where FT grew or stagnated
Moberg 2005	Moberg 2008	Fairtrade	Material related to: social premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation; markets	Ethnographic.	• Environmental criteria increased wage and labour costs and reduced net income • Social premium results in material benefits, but FT net price does not amount to much • FT market is limited; supply of bananas from Windward Islands is greater than demand

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Moore 2010	n/a	Fairtrade	Material related to: monitoring and auditing; social premium use	None	Material related to: cooperative management/performance; markets	Non-ethnographic. Participant observation and semi-structured interviews.	• Historical unbalances power relations between groups can affect the implementation of joint bodies in FT plantations and the use of premium • Audits tend to find visible violations, such as health and safety violations, but are inadequate to identify more complex and less obvious issues • Indian laws and regulation go above and beyond FT guarantees • Predisposition of plantation management to promote and protect workers' rights is more effective than FT certification
Moyo and Mugabe 2014	n/a	Fairtrade	Material related to: implementation costs of certification programmes; social premium use	None	Material related to: markets; production cost for certified goods	Non-ethnographic. Interviews and focus group discussions.	• Social dividends can be a positive force in the community • Certification costs can be a burden on emerging farmers • FT can reduce farmers' marketing flexibility if FT prices drop because the farmers can't easily/quickly shift their produce to another market

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Naylor 2014	n/a	Fairtrade, organic	Material related to: monitoring and auditing; financial premium use; social premium use	None	Material related to: cooperative management/performance; social infrastructure; markets; production cost for certified goods	Ethnographic.	• Development projects funded by FT premium can be problematic in cases of politically polarised communities like Chiapas. • The importance of cooperative ownership and the quality of the coop-buyer relationship for effective FT programs • FT programmes have filled in a vacuum left by the collapse of public coffee institutions in Mexico
Nelson et al 2002	n/a	Fairtrade	Material related to: training and new practices; certification related services	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; markets	Non-ethnographic. Interviews.	• Wage workers involved in brazil nut production have not been considered by ethical trade schemes and receive no benefits • There is a wealth and gender bias in benefits from ethical trade • Candela is offering credit in better terms than others on the market and has been more transparent in its pricing, but it is not performing in other aspects such as creating trust with collectors, or being a local quality standard setter.

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Nelson and Martin 2013a	n/a	Fairtrade	Material related to: certification related services; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Non-ethnographic. Survey.	• No significant differences between certified and non-certified producers were found in several aspects, i,e. access to credit and extension services, gender division of labour, and hire labour conditions etc. besides training • The personal relationship with the person who purchases the product is more important than the certification or the coop • 30% of certified producers did not farm their own farms but used labour
Nelson and Martin 2013 b (Kenya; Wage workers)	n/a	Fairtrade, Rainforest Alliance	Material related to: training and new practices; costs of certification programmes	Material related to: distribution of benefits & investments	Material related to: markets	Non-ethnographic. Interviews.	• GAP improved production but stricter quality criteria lead to an overall decrease in volumes • Labour standards, particularly regarding overtime, can be conflicting as workers end up with less income due to overtime restrictions
Nelson and Martin 2013b (Kenya; Producers)	n/a	Fairtrade, Rainforest Alliance	Material related to: costs of certification programmes; financial premium use	Material related to: distribution of benefits & investments; elite capture	None	Non-ethnographic. Interviews.	• Floor price is much lower than market price and therefore this mechanism does not have an impact

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Nelson and Martin 2013b (India)	n/a	Fairtrade, Rainforest Alliance	Material related to: certification related services; social premium use	Material related to: distribution of benefits & investments	None	Non-ethnographic. Interviews.	• Existing national legislation covers and goes beyond FT labour standards and no differences were noted between certified and non-certified estates. • Possible negative impact on income due to overtime restrictions. • Workers in certified estates appear to have better access to credit
Nelson and Martin 2013b (Ecuador)	n/a	Fairtrade, Rainforest Alliance	None	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Non-ethnographic. Interviews.	• Certified coop has access to better facilities, and perhaps because of this they are able to sell all of their production to Fairtrade buyers.
Nelson and Martin 2013b (Ghana)	n/a	Fairtrade, Rainforest Alliance	None	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Non-ethnographic. Interviews.	• Caretakers, the backbone of Ghana's cocoa production, are not able to join Kuapa Kokoo and join FT
Ouma 2010	n/a	GlobalGAP	Material related to: training and new practices; implementation costs of certification programmes; financial premium use	None	Material related to: markets	Non-ethnographic. Field research, interviews, questionnaires, and literature reviews.	• Highlights issues like cost of implementation, "backchannels" through which uncertified farmers could continue to sell produce, and revisions of the protocol towards more locally adapted solutions

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Pollack 2006	n/a	Fairtrade, organic	None	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation	Ethnographic.	• Women's only groups in FT coops facilitate women's participation (including in in decision making) and seem to be able to overcome barriers from lack of landownership or not being able to officially sell the crop • Machismo and women's reproductive labour are barriers to participation • Government funding is important to keep up the women's projects
Pongratz-Chander 2007	n/a	Fairtrade, organic	Material related to: financial premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation; markets	Non-ethnographic. Observation.	• Largest 1st level coops are being assigned greater quotas for the alternative market. • FT and aid agencies give priority to stable, well established groups, excluding many of the poorest and most marginal of producers • Women's only groups and projects are described as a way to overcome barriers in women's participation

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Quaedvlieg et al 2014	n/a	Fairtrade, organic	Material related to: training and new practices; implementation costs of certification programmes	None	Material related to: participants' reception/motivation; markets	Non-ethnographic. Semi-structured interviews.	• Certification is implemented on the initiative of NGOs rather than associations. • Small associations in combination with limited demand for FT meant they can't deal with certification costs and depend on NGOs. • Cultural differences are a barrier to cooperation between producers
Raynolds 2012	Raynolds 2014	Fairtrade	Material related to: training and new practices; certification related services implementation costs of certification programmes; social premium use	Material related to: uneven adoption of standard	Material related to: markets	Non-ethnographic. Interviews.	• Meeting FLO's environmental and health and safety standards is difficult and costly • Certified companies are industry leaders and have helped demonstrate best practices • FT standards go beyond country labour laws

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Raynolds 2014	Raynolds 2012	Fairtrade	None	None	Material related to: participants' reception/motivation	Non-ethnographic. Interviews.	• Workers need to have power to ensure standards are respected when it comes to anti-discriminatory procedures • FT promoted workers' committees in the vacuum of lack of trade unions, but they can be a poor substitute as they can be vulnerable to management pressure
Riisgaard et al 2009 (Uganda, Coffee)	n/a	Fairtrade, organic, UTZ	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation	Non-ethnographic. Focus group discussions and quantitative household survey.	• It is impossible to improve quality without training women since women do most of the work in the coffee gardens, but training women didn't happen automatically and had to be insisted upon • Being paid at point of sale made real investments possible • Overall positive perceptions of certification were recorded

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Riisgaard et al 2009 (Uganda, Tea)	n/a	Fairtrade	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation	Non-ethnographic. Focus group discussions and quantitative household survey.	• Prompt payment and possibility of urgent advances provided more security and made it easier to send children to school • Since inception of scheme, women had been trained in participating more in the activities of the coop and participated more in committees and councils • Overall positive perceptions of certification were recorded
Riisgaard et al 2009 (Kenya, Tea)	n/a	Fairtrade	Material related to: financial premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation	Non-ethnographic. Focus group discussions and quantitative household survey.	• An increase in women on committees and councils was seen • Overall positive perceptions of certification were recorded
Riisgaard et al 2009 (Kenya; Coffee)	n/a	UTZ	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation; production cost for certified goods	Non-ethnographic. Focus group discussions and quantitative household survey.	• Donor coverage of initial and recurring certification costs has been essential to success of the programme • Assured sales and related ability to sell in bulk and receive cash lump sum made biggest improvement in situation • Overall positive perceptions of certification were recorded

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Riisgaard et al 2009 (Ethiopia)	n/a	Rainforest Alliance	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation; production cost for certified goods	Non-ethnographic. Focus group discussions and quantitative household survey.	• The exporter and an international trading company initially paid for the certification but decided not to re-certify after a couple of years due to lack of premium • Overall positive perceptions of certification were recorded
van Rijn et al 2016 (Ghana)	n/a	Fairtrade	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: social infrastructure	Non-ethnographic. Interviews, survey, gaming sessions.	• Not all workers were aware of their rights and/or the facilities available to them • There were still very few women in supervisory positions, but this may be more due to the fact that women were more likely to be illiterate and have lower levels of education, rather than explicitly due to gender bias
van Rijn et al 2016 (Colombia)	n/a	Fairtrade	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	None	Non-ethnographic. Interviews, survey, gaming sessions.	• FT workers received substantially more in-kind benefits than non-FT workers • Female wage workers on FT plantations reported a negative change in confidence of job continuation and in-kind benefits, feel less job secure

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
van Rijn et al 2016 (Dominican Republic)	n/a	Fairtrade	Material related to: certification related services; financial premium use	Material related to: distribution of benefits & investments	Material related to: social infrastructure	Non-ethnographic. Interviews, survey, gaming sessions.	• FT workers received more in-kind benefits than non-FT workers and had better housing and food security • Migrant workers from Haiti would like to be able to spend premium benefits back home (eg for education for children, or house repairs) but are not allowed to
Romanoff 2010	n/a	Rainforest Alliance	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	None	None	Non-ethnographic. Interviews.	• An important facilitator is assistance with cost and training: in this study, USAID covered all certification costs and trainings for the farmers, and found that this was a huge benefit in entry to the certification

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Ronchi 2002	n/a	Fairtrade, organic	Material related to: training and new practices; certification related services; financial premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; markets	Non-ethnographic. Interviews.	• Training at the 2nd coop level is helpful for dealing with importers • FT secure price allows long term investments in infrastructure and income diversification programmes • Participation in FT enhances inter-coop cooperation and cooperative lobbying with public institutions in favour of small holders • Many women were members on paper only in order for the family, as a unit, to access greater credit from the co-op
Rotter 2011	n/a	Fairtrade, organic	Material related to: financial premium use	Material related to: uneven adoption of standard	Material related to: governmental legislation, incentives, and/or regulation; social infrastructure; markets	Ethnographic.	• Payment delays are a barrier for smaller producers who cannot afford to wait, even if the price is better. As a result, bigger producers who have enough coffee both for the coop and the coyote are the ones who are able to participate • External initial capital may be necessary to kick start similar projects

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Roy and Thorat 2008	n/a	GlobalGAP	Material related to: participant's targeting and self-selection; implementation costs of certification programmes	Material related to: None	Material related to: effective/non-effective adoption; production cost for certified goods	Non-ethnographic. Field surveys with questionnaire-based interviews.	• The presence of an interested marketing partner provides key support in implementation and certification, especially for small-medium farmers • Participation in Eurepgap/GlobalGAP and the export market leads to higher prices and net revenue despite higher production costs (to meet requirements) • No bias against smallholders was found
Rueda and Lambin 2013	n/a	Rainforest Alliance	Material related to: social premium use	Material related to: uneven adoption of standard	Material related to: participants' reception/motivation; effective/non-effective adoption	Non-ethnographic. Interviews, household survey.	• High premium is a motivation for joining the certification programme, while access to networks, information, technology, and skills are motivation for staying in • The involvement of the Colombian Coffee Growers' Federation made certification accessible for small and large holders

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Ryan 2011	n/a	Fairtrade	Material related to: monitoring and auditing; financial premium use; social premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation; governmental legislation, incentives, and/or regulation; markets	Ethnographic.	• FT floor prices were unlikely to make any real difference to farmers as farm gate prices were likely to increase due to production shortages and political pressure • FT has very limited capacity to bring any real change/improvements due to its limited market share and the strong role the Ghanaian government plays in cocoa production in Ghana
Said-Allsopp and Tallontire 2014	n/a	Fairtrade	Material related to: training and new practices; participant's targeting and self-selection; monitoring and auditing; social premium use	Material related to: distribution of benefits & investments; elite capture	Material related to: cooperative management/performance	Non-ethnographic. Focus group discussions.	• While the Joint Body seemed effective in empowering women, in reality it was a prior institution, "Gender Committees", instituted outside FT interventions, which produced the desired empowerment effects • The JB may actually partly offset empowerment given that it is dominated by male managers and workers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Schelly 2011	n/a	Fairtrade	Material related to: training and new practices; participant's targeting and self-selection; certification related services; monitoring and auditing	Material related to: uneven adoption of standard	Material related to: markets	Non-ethnographic. Semi-structured interviews.	• The worker's committee has created a valuable space of communication between workers and management but has very limited power to deal with real issues, such as pressure from quota system or inadequate compensation • Management manipulates audits to give the best impression • Workers fear losing FT certification and therefore are hesitant to speak out in audits
Schoonhoven-Speijer 2012	n/a	UTZ	Material related to: training and new practices; participant's targeting and self-selection; certification related services	Other	Material related to: cooperative management/performance	Non-ethnographic.	• Members of both certified and uncertified coops had trust in their coops • Most common reason for choosing a cooperative was location. This may in part be due to Kenyan regulations: coffee farmers must sell through a local coop, and berries must be processed within 24 hours of harvest

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Schuster and Maertens 2013	n/a	Various	None	None	Material related to: markets	Non-ethnographic. Secondary datasets and survey.	• Certification for high-level production standards (eg GlobalGap) changes export companies' sourcing strategies and significantly reduces the share of produce sourced from external suppliers in general and from small-scale suppliers in particular, while certification for low-level production standards processing standards don't • GlobalGap in particular significantly decreases external sourcing and sourcing from small farmers • Reasons given by companies include need to assure a certain quality and produce a continuous export flow

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Sen 2009	n/a	Fairtrade, organic	Material related to: monitoring and auditing	Material related to: elite capture	Material related to: cooperative management/performance	Ethnographic.	• FT Joint Bodies failed to become a space where workers can express their demands, • While the women only sub-group of the FT coop on did provide a valid space for women to materialise their plans, there were barriers from the male dominated coop and the FT inspector • FT inspectors failed to understand the reality and made inadequate "empowerment" suggestions that are rejected by women. • Middlemen try to influence use of FT premium to their own interests, opposing use of FT premium for a women's better market access that would harm their business

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Setrini 2011	n/a	Fairtrade, organic	Material related to: training and new practices; participant's targeting and self-selection; monitoring and auditing; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Ethnographic.	• Industry limited engagement with FT due to FLO's lack of resources • Barriers included coop corrupted leadership in combination with cultural context of authoritarianism and fear, monopsonistic environment, and producers' fear and passive attitudes • Processing factories implemented quotas and manipulated access to markets prioritising its own production and larger producers • Organic became compulsory, as it was the only way to sell

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Schreck 2002	Getz and Shreck 2006	Fairtrade, organic	Material related to: training and new practices; certification related services; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: cooperative management/performance; participants' reception/motivation; markets	Ethnographic.	• Quality requirements make costs of certified production higher and small-scale farmers cannot deal with them • Lack of knowledge and understanding about the FT market was creating disappointment and mistrust among farmers • FLO has limited power to enforce premium payments for defaulting importers • Growers did not know when their product is sold at the FT price or not
Silva-Castaneda 2012	n/a	RSPO	None	None	Material related to: governmental legislation, incentives, and/or regulation	Non-ethnographic. Interviews, observation.	• Problems with implementing RSPO with regards to land disputes include too many parties involved in the certification process, (companies, NGOs, etc), and not enough evidence • Evidence usually means "documents", which companies tend to have and smallholders and villagers don't, which puts companies at an advantage

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Smith et al 2004	n/a	EurepGAP	None	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance	Non-ethnographic. Interviews, focus groups.	• Coffee quality and location can be a barrier to access of FT markets • Relationship with a buyer was key to the formation of a new coop. • Very limited percentage of production was actually sold to alternative markets
Smith 2007	n/a	Fairtrade, organic	Material related to: financial premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance	Ethnographic.	• Discrimination against seasonal/non-permanent workers and women • Poor implementation of codes especially in regards to labour
Smith 2010 (Ghana)	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: markets	Non-ethnographic. Interviews.	• Enforcement of labour standards provided indefinite written contracts to plantation workers which enabled them to have the benefits of formal employment • Migrants were almost systematically disadvantaged in allocation of FT premium • Quality and standards required for Fairtrade markets acted as a disincentive to participate • Social premium was being used to cover core business expenses

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Smith 2010 (Ecuador)	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: markets	Non-ethnographic. Interviews.	• Inclusion of larger producers was helping SPOs to reach required volumes for export and keep providing affordable services to the small producers- although in practice controversial. • Enforcement of labour standards provided indefinite written contracts to plantation workers which enabled them to have the benefits of formal employment • Migrants were almost systematically disadvantaged in allocation of FT premium • Quality and standards required for Fairtrade markets acted as a disincentive to participate

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Smith 2010 (Various)	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: markets	Non-ethnographic. Interviews.	• Inclusion of larger producers was helping SPOs to reach required volumes for export and keep providing affordable services to the small producers- although in practice controversial. • Enforcement of labour standards provided indefinite written contracts to plantation workers which enabled them to have the benefits of formal employment • Migrants were almost systematically disadvantaged in allocation of FT premium • Quality and standards required for Fairtrade markets acted as a disincentive to participate

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Smith 2010 (Windward Islands)	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: markets	Non-ethnographic. Interviews.	• Inclusion of larger producers was helping SPOs to reach required volumes for export and keep providing affordable services to the small producers- although in practice controversial. • Enforcement of labour standards provided indefinite written contracts to plantation workers which enabled them to have the benefits of formal employment • Migrants were almost systematically disadvantaged in allocation of FT premium • Quality and standards required for Fairtrade markets acted as a disincentive to participate

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Smith 2010 (Dominican Republic)	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: markets	Non-ethnographic. Interviews.	• Inclusion of larger producers was helping SPOs to reach required volumes for export and keep providing affordable services to the small producers- although in practice controversial. • Enforcement of labour standards provided indefinite written contracts to plantation workers which enabled them to have the benefits of formal employment • Migrants were almost systematically disadvantaged in allocation of FT premium • Quality and standards required for Fairtrade markets acted as a disincentive to participate

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Staib 2012	n/a	Fairtrade, organic	Material related to: certification related services; monitoring and auditing; implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments; elite capture	Material related to: cooperative management/performance; participants' reception/motivation; markets	Ethnographic.	• Smaller producers/coops get less attention • Social premium programs are instrumentalised to serve other purposes besides those of the direct beneficiaries • Cooperatives may overestimate members production in order to access greater loans, and make up for the difference with non-certified production
Staricco and Ponte 2015	n/a	Fairtrade	Material related to: certification related services	Material related to: uneven adoption of standard	Material related to: markets	Non-ethnographic. Semi-structured interviews.	• Fair Trade requires minimum wage to be legal, but does not consider its actual purchasing power, which combined with the overtime restrictions lead to a lack of integrative income. • Wine workers hired a nutritionist to "dress" premium with a health concern and be able to increase their income in kind • FT in Argentina wine has mainly benefited the large producers and has done little to incorporate the smaller and marginalised producers

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Stathers and Gathuthi 2013	n/a	Fairtrade, Rainforest Alliance	Material related to: training and new practices; certification related services; implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; participants' reception/motivation; markets	Non-ethnographic. Formal questionnaires and qualitative checklists, interviews, focus group discussions.	• Both certifications raised awareness in farmers about the importance of joint planning and decision-making and improved labour rights and conditions • Certified organisations generally did not keep track of certification costs and felt they were worth it • Certification has increased transparency and communication • Women were much less informed about certification and its related benefits than men
Stenn 2015	n/a	Fairtrade	Material related to: training and new practices; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: participants' reception/motivation	Ethnographic.	• Barriers to female participation include lack of confidence, lack of time to leave the farm/children, and more blatant discrimination such as failure to communicate meeting times, schedule them at times women can attend, or cover travel • Overall, participants were split between support for and mistrust of FT

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Subervie and Vagneron 2013	n/a	GlobalGAP	Material related to: training and new practices; participant's targeting and self-selection	Material related to: distribution of benefits & investments	None	Non-ethnographic. Surveys, questionnaires.	• Location and mobility were barriers; only producers who could travel and sell direct in the main market town received higher prices for certified goods • Other barriers to entry included education and volume of sales
Sutton 2014	n/a	Fairtrade	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; markets	Non-ethnographic. Interviews.	• Payment delays in combination with lack of transparency and good communication is a barrier to effectiveness • Use of premium is decided at the coop headquarters and not by participatory decision making in assemblies • Only land owners are eligible for coop membership eligibility, excluding female farmers • it appears FT does not really alter the governance conditions and therefore its outcomes will always be dependent on the specificities of the coop union in which it operates

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Terstappen 2010	n/a	Fairtrade, organic	Material related to: training and new practices; implementation costs of certification programmes; financial premium use; social premium use	Material related to: uneven adoption of standard	Material related to: cooperative management/performance; markets	Ethnographic.	• FT price was not enough given the work involved in certified production • Certification costs were expensive • Certification was an obligation, not a choice, for market access
Trauger 2014	n/a	Fairtrade, organic	Material related to: training and new practices; monitoring and auditing; implementation costs of certification programmes; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; markets	Ethnographic.	• No difference between FT and non FT farms for wage workers in working conditions • Inequities experienced by Haitian wage workers in FT smallholder farms was reported • Workers in plantations benefit more from FT than those working in smallholder farms • Audits seem to be much stricter for plantations than for smallholders, who get a notice before getting audited

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
TWIN 2012 (Côte d'Ivoire)	n/a	Fairtrade	Material related to: training and new practices; certification related services; financial premium use	None	Material related to: cooperative management/performance; governmental legislation, incentives, and/or regulation; markets	Non-ethnographic. Focus group discussions and SSIs.	• There is a lot of variability and differentiation of outcomes depending on the specific set up Price, and ability to influence it, seemed to play a less significant role in relative returns than productivity • Pre-finance is a key element in determining whether co-operatives are able to operate on the same playing field as better resourced traders and processors • Overall the scarce influence on price and trade dynamics suggests that the potential for empowerment in current market conditions is limited
TWIN 2012 (Kenya)	n/a	Fairtrade	Material related to: training and new practices	None	Material related to: cooperative management/performance	Non-ethnographic. Focus group discussions and SSIs.	• There is a lot of variability and differentiation of outcomes depending on the specific set up Price, and ability to influence it, seemed to play a less significant role in relative returns than productivity • Timing of returns was key in tea and groundnuts • Overall the scarce influence on price and trade dynamics suggests that the potential for empowerment in current market conditions is limited

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
TWIN 2012 (Malawi)	n/a	Fairtrade	Material related to: training and new practices	None	Material related to: cooperative management/performance	Non-ethnographic. Focus group discussions and SSIs.	• There is a lot of variability and differentiation of outcomes depending on the specific set up Price, and ability to influence it, seemed to play a less significant role in relative returns than productivity • Timing of returns was key in tea and groundnuts • Overall the scarce influence on price and trade dynamics suggests that the potential for empowerment in current market conditions is limited
TWIN 2013	n/a	Fairtrade	None	Material related to: distribution of benefits & investments; uneven adoption of standard	Material related to: social infrastructure	Non-ethnographic. Interviews, group discussions.	• POs and certification organisations must work with an awareness of cultural norms and prejudices when working towards participation of women and gender equality • Barriers to women's full participation include land ownership, cultural norms, and confidence

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Valkila and Nygren 2008	Valkila 2009	Fairtrade, organic	Material related to: training and new practices; certification related services; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; markets	Non-ethnographic. Semi-structured interviews.	• FT premium does not add to the organic premium • FT coops offer worse credit terms than big private companies • There was pressure to be organic in order to access FT • Large part of premiums are used to improve cooperative infrastructure and to pay for organic and Fair Trade certifications
Valkila 2009	Valkila and Nygren 2008	Fairtrade, organic	Material related to: certification related services; implementation costs of certification programmes; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: markets	Non-ethnographic. Semi-structured interviews.	• Price differential vanishes when world market prices are on the rise and equal the FT ones • In times of low mainstream prices, coops can demand higher quality coffee in exchange of the premium. However when FT prices equal the market prices, farmers prefer the mainstream market as it pays faster, selling points are closer and it's less demanding in terms of quality. • Pre-financing seems not to have enabled FT coops to provide credit in favourable terms

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Vásquez-León 2010	n/a	Fairtrade, organic	Material related to: financial premium use; social premium use	None	Material related to: markets	Ethnographic.	• Monopsony of FT processor results in lower prices and abuses of the premium • Lack of government support is a barrier to growth
Waarts et al 2012	n/a	Rainforest Alliance	Material related to: training and new practices	None	Material related to: participants' reception/motivation; production cost for certified goods	Non-ethnographic. Quantitative survey, focus group discussions.	• Certification costs can be challenging for small farmers • Training led to greater knowledge and improved production and income
Waarts et al 2014	n/a	UTZ	Material related to: training and new practices; participant's targeting and self-selection	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; participants' reception/motivation	Non-ethnographic. Interviews, focus groups, quantitative survey.	• Highlighted the success of training with regards to positive implementation of sustainable practices, increases in net income, and farmer perception • Found that continual support for training and for lead farmers is integral to programme success
Waarts et al 2016	n/a	UTZ	Material related to: training and new practices; certification related services	Material related to: uneven adoption of standard	None	Non-ethnographic. Interviews, survey, focus group discussions.	• Implementation and knowledge of GAP has improved through training • Women were not generally included in the programme, in large part because they were not landowners

Study	Associated papers	Certification scheme(s)	Evidence on implementation dynamics	Evidence on distributional dynamics	Evidence on causal mechanisms and barriers and facilitators	Methodology	Main findings
Walsh 2004	n/a	Fairtrade, organic	Material related to: certification related services; financial premium use; social premium use	Material related to: distribution of benefits & investments	Material related to: cooperative management/performance; markets; production cost for certified goods	Non-ethnographic. Semi-structured interviews.	• Fair Trade buyers don't always provide pre- harvest financing, yet relationships with Fair Trade importers still play an important role in securing credit. FT contracts are used as a collateral for accessing credit • Barriers to certification include paperwork for organic production, low prices and no premium during transition period
Wilson 2010	n/a	Fairtrade	Material related to: certification related services; financial premium use	Material related to: elite capture	Material related to: cooperative management/performance; participants' reception/motivation; markets	Ethnographic.	• In times of acute crisis, FT process & credit schemes may not be enough, as farmers enter in cycle of indebtedness that is hard to escape • Credit in exchange of future harvest can motivate producers to increase production but also bear risks if production fluctuates and does not meet the expected volumes • FT coop credit schemes offer better terms (lower interests and more flexible repayment rules) but still they do not offer sufficient to sustain production in times of crisis

Annex C: Detailled risk of bias assessments

Risk of bias assessment for review question 1 studies

Study	Selection and confound-ing	Group equivalence	Motivation bias	Spill-over effects	Selective reporting of outcome	Selective analysis	Other sources of bias	Final score
Anteneh et al, 2014	UNCLEAR/NO	UNCLEAR/NO	YES	UNCLEAR	YES	UNCLEAR	YES	HIGH
Asfaw et al, 2010	UNCLEAR	UNCLEAR	YES	YES	YES	YES	NO	HIGH
Balineau, 2012	UNCLEAR	UNCLEAR	YES	NO	YES	UNCLEAR	YES	HIGH
Barham and Weber, 2012	NO	NO	UNCLEAR	UNCLEAR/NO	YES	NO	YES	CRITICAL
Becchetti and Gianfreda, 2008	NO	NO	YES	UNCLEAR	YES	UNCLEAR	NO	HIGH
Becchetti et al, 2008	NO	NO	YES	UNCLEAR	YES	YES	NO	HIGH
Becchetti et al, 2011	NO	UNCLEAR	YES	UNCLEAR	YES	UNCLEAR	NO	HIGH
Cepeda et al, 2013	NO	NO	YES	NO	YES	YES	UNCLEAR	CRITICAL
Chiputwa and Qaim, 2014	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Colen et al, 2012	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Bennett et al, 2012	UNCLEAR	UNCLEAR	YES	YES	YES	YES	UNCLEAR	MODERATE
Cramer et al, 2014 (Ethiopia, coffee)	YES/UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Cramer et al, 2014 (Ethiopia, flowers)	YES/UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Cramer et al, 2014 (Uganda, coffee)	YES/UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Cramer et al, 2014 (Uganda, tea)	YES/UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Dragusanu, 2014	NO	NO	YES	YES	YES	YES	UNCLEAR	HIGH

Ehlert et al, 2014	NO	NO	YES	UNCLEAR	YES	UNCLEAR	UNCLEAR	HIGH
Fort and Ruben, 2009	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
García et al, 2014	UNCLEAR/NO	UNCLEAR/YES	YES	YES	YES	UNCLEAR	YES	MODERATE
Jena et al, 2012	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	UNCLEAR	NO	HIGH
Kamau et al, 2010	NO	NO	YES	UNCLEAR	YES	UNCLEAR	UNCLEAR	HIGH
Kuit et al, 2016 (Uganda)	UNCLEAR	NO	YES	YES	YES	UNCLEAR	UNCLEAR	HIGH
Kuit et al, 2016 (Vietnam)	UNCLEAR	NO	YES	YES	YES	UNCLEAR	UNCLEAR	HIGH
Minten et al, 2015	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	UNCLEAR	YES	MODERATE
Mueller and Theuvsen, 2015	UNCLEAR	UNCLEAR/YES	YES	UNCLEAR	YES	YES	YES	MODERATE
Nelson et al, 2013	NO	NO	YES	NO	YES	NO	NO	CRITICAL
Parvathi and Waibel, 2016	NO	UNCLEAR/NO	YES	UNCLEAR	YES	UNCLEAR	NO	HIGH
Riisgaard et al, 2009 (FT)	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	UNCLEAR	UNCLEAR	MODERATE
Riisgaard et al, 2009 (UTZ)	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	UNCLEAR	UNCLEAR	MODERATE
Roy and Thorat, 2008	UNCLEAR/NO	UNCLEAR/NO	YES	UNCLEAR	YES	UNCLEAR	NO	HIGH
Ruben and Fort, 2012	UNCLEAR	UNCLEAR	YES	UNCLEAR	YES	YES	UNCLEAR	MODERATE
Ruben and Zúñiga-Arias, 2011	NOT POSSIBLE TO ASSESS	NOT POSSIBLE TO ASSESS	YES	YES	YES	YES	NO	UNCLEAR
Ruben et al, 2014	UNCLEAR/NO	UNCLEAR/NO	YES	UNCLEAR	YES	YES	UNCLEAR	HIGH
Schuster and Maertens, 2014	UNCLEAR	YES/UNCLEAR	YES	YES	YES	YES	YES	LOW
Stathers and Gathuthi, 2013	NO	NO	YES	NO	YES	YES	NO	CRITICAL
Subervie and Vagneron, 2013	UNCLEAR/NO	UNCLEAR	YES	YES	YES	YES	NO	HIGH

van Rijn et al, 2016 (Ghana)	NO	NO	YES	UNCLEAR	YES	UNCLEAR	YES	CRITICAL[98]
van Rijn et al, 2016 (Colombia)	UNCLEAR/NO	NO	YES	UNCLEAR	YES	UNCLEAR	YES	HIGH
van Rijn et al, 2016 (Dominican Republic)	UNCLEAR/NO	NO	YES	UNCLEAR	YES	UNCLEAR	YES	HIGH
van Rijsbergen et al, 2016	UNCLEAR/YES	YES	YES	UNCLEAR	YES	UNCLEAR	UNCLEAR/YES	LOW
Waarts et al, 2012	YES	UNCLEAR/YES	YES	UNCLEAR/YES	YES	YES	UNCLEAR	HIGH
Waarts et al, 2016	UNCLEAR/NO	UNCLEAR/NO	UNCLEAR	UNCLEAR	YES	NO	NO	HIGH
Weber, 2011	UNCLEAR	UNCLEAR	YES	YES	YES	YES	YES	MODERATE

[98] This study was rated as critical despite not fulfilling our normal criteria due to the complete absence of a meaningful control group.

Risk of bias assessment for review question 2 studies

Study	Research question	Research Approach	Research Context	Researcher's role	Sampling methods	Site selection	Data Collection	Analysis	Evidence	Triangulation
Abarca-Orozco 2015	Yes	Not clear	Yes	Not clear	Yes	Yes	Yes	Yes	Yes	Not reported
Aidenvironment 2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not reported
Amekawa 2001	Yes	Not clear	Yes	Not clear	Not clear	Not reported	Yes	Not reported	Yes	Not reported
Arce 2009	Not clear	Yes	Yes	No	No	No	Yes	No	Yes	No
Asfaw et al 2009	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not reported
Babin 2012	Yes	Yes	Yes	Not clear	N/A	Yes	Not clear	Not clear	Yes	Yes
Bacon 2005	Yes	Not clear	Yes	Not clear	Yes	Not clear	Yes	Not reported	Not clear	Yes
Bagama et al 2014	Yes	Yes	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Bakker 2014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Balineau 2011	Yes	Yes	Yes	Not reported	Yes	Not reported	Yes	Yes	Yes	Not reported
Barham and Weber 2012 (Mexico)	Yes	Yes	Yes	No	Not clear	Not reported	Yes	Yes	Yes	Not reported
Barham and Weber 2012 (Peru)	Yes	Yes	Yes	Not reported	Not clear	Not reported	Yes	Yes	Yes	Not reported
Beall 2012	Not clear	Not reported	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Bergeron 2010	Not clear	Not clear	Yes	Not reported	Yes	Not reported	Yes	Yes	Yes	Not clear
Besky 2014	Yes	Yes	Yes	Yes	Not reported	Yes	Yes	Not reported	Yes	Not reported
Beuchelt 2009	Not clear	Not clear	Not clear	Not reported	Yes	Not reported	Yes	Not reported	No	Not reported
Bonanno and Cavalcanti 2012	Yes	Yes	Yes	Not reported	Not reported	Not reported	Yes	Yes	Yes	Not clear
Brown 2012	Yes	Yes	Yes	No	No	Not reported	Yes	Not reported	Not clear	No
Carimentrand and Ballet 2010	Not clear	Not reported	Yes	Not reported	Not clear	Not reported	Not clear	Not reported	No	Not reported

CESU 2012 (Ecuador)	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
CESU 2012 (Kenya)	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
CESU 2012 (Ghana)	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
CESU 2012 (Peru; Coffee)	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
CESU 2012 (Peru; Banana)	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
CESU 2012 (India)	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
Cepeda et al 2013	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes	Yes
Chiputwa et al 2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Cofre et al 2012	Yes	Not clear	Not clear	No	Yes	Not clear	Yes	Yes	Not clear	No
Cramer et al 2014a (Ethiopia; Coffee)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cramer et al 2014a (Ethiopia; Flowers)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cramer et al 2014a (Uganda; Coffee)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cramer et al 2014a (Uganda; Tea)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cramer et al 2014b (Uganda)	Yes	Yes	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Cramer et al 2014b (Ethiopia)	Yes	Yes	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Dannenberg and Nduru 2013	Yes	Yes	Yes	Not clear	Yes	Not clear	Yes	Not clear	Yes	Not reported
Dolan 2010	Yes	Yes	Yes	Not clear	Yes	Not reported	Yes	Not clear	Yes	Yes
Donovan and Poole 2014	Not clear	Not reported	Yes	Not reported	Yes	Not reported	Yes	Yes	Yes	Not reported
Donovan and Poole 2014	Not clear	Not clear	Yes	Not reported	Not clear	Not clear	Yes	Not reported	Yes	Not reported
Dowdall 2012	Yes	Yes	Yes	No	Yes	Yes	Yes	Not clear	Yes	Yes
Dragusanu and Nunn 2014	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No

Ellery 2010	Yes	Yes	Yes	Not reported	Yes	Not reported	Yes	Yes	Yes	Yes
Fairtrade Foundation 2013 (Tea)	Not clear	Yes	Yes	No	No	Not clear	Yes	Not reported	Not clear	Yes
Fairtrade Foundation (Groundnuts) 2013	Not clear	Yes	Yes	No	No	Yes	Yes	Not reported	Not clear	Yes
Fairtrade Foundation 2013 (Sugar)	Not clear	Yes	Yes	No	Yes	Yes	Yes	Not reported	Not clear	Yes
Fairtrade Foundation 2015 (Bananas)	Yes	Not clear	Yes	No	Not clear	Yes	Not clear	Not reported	Not clear	Not reported
Fairtrade Foundation 2015 (Cotton)	Yes	Not clear	Yes	No	Not clear	Yes	Not clear	Not reported	Not clear	Not reported
Fairtrade Foundation 2015 (Tea)	Yes	Not clear	Yes	No	Not clear	Yes	Not clear	Not reported	Not clear	Not reported
Fayet and Vermeulen 2014	Yes	Yes	Yes	Not reported	Not reported	Not reported	Yes	Not reported	No	Not reported
Francesconi and Ruben 2014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fraser et al 2014	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Not clear
García et al 2014	Yes	Yes	Yes	Not reported	Yes	Yes	Yes	Yes	Yes	Not reported
Getz and Shreck 2006	Not clear	Not clear	Not clear	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported
Gómez-Cardona 2012	Not clear	Not clear	No	No	No	No	Yes	Not reported	Yes	No
Hanson et al 2012	Not clear	Yes	No	Not reported	No	Not reported	Yes	Not reported	Not clear	Not reported
Heller 2010	Not clear	Yes	Yes	Yes	Not reported	Yes	Yes	No	Yes	Not clear
Herman 2010	Yes	Not reported	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Jaffee 2006	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes

Jari et al 2013	Yes	Not reported	Yes	Not reported	Yes	Not reported	Not clear	Yes	Not clear	Not reported
Jena et al 2012	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No
Kariuki 2014	Yes	Not reported	Yes	Not reported	Not reported	Not reported	Yes	Yes	Yes	Not reported
Köhne 2014	Yes	Not reported	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Laroche et al 2012	Not clear	Not reported	Yes	Not reported	Yes	Yes	Yes	Not reported	Not clear	Not reported
Larsen et al 2014	Yes	Yes	Yes	Not clear	Not reported	Not clear	Yes	Not reported	Yes	Not reported
Loconto and Simbua 2010	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes	Not clear
Luetchford 2008	Yes	Yes	Yes	Yes	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Lyall 2014	Not clear	Not clear	Yes	Not clear	Not clear	Yes	Yes	Yes	Yes	No
Lyon 2005	No	Yes	Yes	No	Not reported	Yes	Yes	Not reported	Not clear	Not clear
Lyon 2006	No	Yes	Yes	No	No	No	Yes	No	Yes	Not clear
Lyon 2007	No	Yes	Yes	No	No	No	Yes	No	Not clear	Not clear
Lyon et al 2010	No	Not clear	Not clear	No	Not reported	Not reported	Yes	Not reported	No	Not reported
Makita 2011	Yes	Yes	Yes	Not reported	Not clear	Yes	Yes	Not reported	Yes	Not reported
Makita 2012	Yes	Not reported	Yes	Not reported	Not clear	Yes	Yes	Not clear	Yes	Not reported
Melkeraaen 2009	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Not reported	Yes	Not reported
Mendez 2002	Not clear	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported
Milford 2004	Not clear	Not clear	Yes	No	Not reported	Yes	Yes	Not reported	Not clear	No
Milford 2014	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	No
Minten et al 2015	Yes	Yes	Yes	Not reported	Yes	Yes	Yes	Yes	Yes	Not reported
Moberg 2005	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported

Moberg 2008	Yes	Not reported	Yes	Not clear	Not reported	Not reported	Not clear	Not reported	Yes	Not reported
Moore 2010	Not clear	Not reported	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Moyo and Mugabe 2014	Yes	Yes	Yes	Not reported	Yes	Not reported	Yes	Not reported	Not clear	Not reported
Naylor 2014	Yes	Yes	Yes	Yes	Not clear	Not reported	Yes	Not clear	Yes	Not clear
Nelson et al 2002	Yes	Yes	No	Not reported	Yes	Not clear	Yes	Yes	No	Not reported
Nelson and Martin 2013a	Yes	Yes	No	Not reported	Yes	Not clear	Yes	Yes	No	Not reported
Nelson and Martin 2013 b (Kenya; Wage workers)	Yes	Not clear	Yes	Not reported	Yes	Yes	Yes	Yes	Yes	Not reported
Nelson and Martin 2013b (Kenya; Producers)	Yes	Yes	No	Not reported	Yes	Not clear	Yes	Yes	No	Not reported
Nelson and Martin 2013b (India)	Yes	Yes	No	Not reported	Yes	Not clear	Yes	Yes	No	Not reported
Nelson and Martin 2013b (Ecuador)	Yes	Yes	No	Not reported	Yes	Not clear	Yes	Yes	No	Not reported
Nelson and Martin 2013b (Ghana)	Not clear	Yes	Yes	Not reported	Not clear	Yes	Not clear	Yes	Not clear	Yes
Ouma 2010	Not reported	Yes	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported
Pollack 2006	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes
Pongratz-Chander 2007	Yes	Not reported	Yes	Not reported	Yes	Not clear	Yes	Not reported	Not clear	Not reported
Quaedvlieg et al 2014	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported	Not clear	Not reported
Raynolds 2012	Not clear	Not clear	Yes	No	Not clear	Not clear	Yes	Not reported	Yes	No
Raynolds 2014	Not clear	Yes	Not clear	Not reported	Not reported	Not reported	Not clear	Not clear	Not clear	Not reported
Riisgaard et al 2009 (Uganda; Coffee)	Yes	Yes	Yes	Not reported	Not reported	Yes	Yes	Yes	Yes	Not reported
Riisgaard et al 2009 (Uganda; Tea)	Yes	Yes	Yes	Not reported	Not reported	Yes	Yes	Yes	Yes	Not reported

Riisgaard et al 2009 (Kenya; Tea)	Yes	Yes	Yes	Not reported	Not reported	Yes	Yes	Yes	Yes	Not reported
Riisgaard et al 2009 (Kenya; Coffee)	Yes	Yes	Yes	Not reported	Not reported	Yes	Yes	Yes	Yes	Not reported
Riisgaard et al 2009 (Ethiopia)	Yes	Yes	Yes	Not reported	Not reported	Yes	Yes	Yes	Yes	Not reported
van Rijn et al 2016 (Ghana)	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes	Yes
van Rijn et al 2016 (Colombia)	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes	Yes
van Rijn et al 2016 (Dominican Republic)	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Yes	Yes	Yes
Romanoff 2010	Yes	Not reported	Yes	Not reported	Not clear	Not reported	Not clear	Not reported	Yes	Not reported
Ronchi 2002	No	Not clear	Yes	Not reported	Yes	Not clear	Yes	Not reported	Yes	Not reported
Rotter 1999	Yes	Yes	Yes	Not clear	Not reported	Not reported	Yes	Yes	Yes	Not reported
Roy and Thorat 2008	Yes	Yes	Yes	Not reported	Yes	Not reported	Yes	Yes	Yes	Not reported
Rueda and Lambin 2013	Yes	Yes	Yes	Not reported	Yes	Yes	Yes	Yes	Yes	Not reported
Ryan 2011	Not reported	Not reported	Yes	Not reported	Not reported	Yes	Not reported	Not reported	Yes	Not reported
Said-Allsopp and Tallontire 2014	Yes	Yes	Not clear	No	Not clear	Not clear	Yes	Yes	Yes	Yes
Schelly 2011	Yes	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Not reported
Schoonhoven-Speijer 2012	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Schuster and Maertens 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not reported
Sen 2009	Yes	Yes	Yes	Yes	Yes	Not reported	Yes	Not reported	Yes	Not reported
Setrini 2011	Yes	Yes	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported
Shreck 2002	Yes	Yes	Yes	Not reported	Yes	Yes	Yes	Not reported	Not clear	Yes

Silva-Castaneda 2012	Not clear	Yes	Yes	Yes	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Smith 2007	No	No	Yes	Not reported	Not reported	Not reported	Not clear	Not reported	Yes	Not reported
Smith 2010 (Ghana)	Not clear	Not clear	Yes	Not reported	Not clear	Yes	Yes	No	Not clear	Not reported
Smith 2010 (Ecuador)	Not clear	Not clear	Yes	Not reported	Not clear	Yes	Yes	No	Not clear	Not reported
Smith 2010 (Various)	Not clear	Not clear	Yes	Not reported	Not clear	Yes	Yes	No	Not clear	Not reported
Smith 2010 (Windward Islands)	Not clear	Not clear	Yes	Not reported	Not clear	Yes	Yes	No	Not clear	Not reported
Smith 2010 (Dominican Republic)	Not clear	Not clear	Yes	Not reported	Not clear	Yes	Yes	No	Not clear	Not reported
Smith et al 2004	Yes	Yes	Yes	No	Yes	Yes	Yes	Not reported	Yes	Not reported
Staib 2012	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not reported	Yes	Not reported
Staricco and Ponte 2015	Not clear	Yes	Yes	Not reported	Not reported	Not reported	Yes	Yes	Yes	Not reported
Stathers and Gathuthi 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not reported
Stenn 2015	Yes	Yes	Yes	Not clear	Not reported	Yes	Yes	Not reported	Yes	Not reported
Subervie and Vagneron 2013	Yes	Not reported	Yes	Not reported	Not clear	Not clear	Not clear	Yes	Yes	Not reported
Sutton 2014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not clear	Yes
Terstappen 2010	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not reported
Trauger 2014	Not clear	Yes	Yes	Not reported	Not reported	Not reported	Yes	Yes	Yes	Not reported
TWIN 2012 (Côte d'Ivoire)	Not clear	Not clear	Yes	No	Not reported	Not clear	Yes	Yes	Not clear	Not clear
TWIN 2012 (Kenya)	Not clear	Not clear	Yes	No	Not reported	Not clear	Yes	Yes	Not clear	Not reported
TWIN 2012 (Malawi)	Not clear	Not clear	Yes	No	Not reported	Not clear	Yes	Yes	Not clear	Not reported

TWIN 2013	Yes	Yes	Yes	Not reported	Not reported	Not reported	Yes	Not reported	Yes	Not reported
Valkila 2009	Not clear	Not reported	Yes	Not reported	Not clear	Not reported	Yes	Not reported	Yes	Not reported
Valkila and Nygren 2008	Not clear	Not reported	Yes	Not reported	Not clear	No	Yes	No	Yes	Not clear
Vásquez-León 2010	Not clear	Not reported	Yes	Not reported	Not reported	Not reported	Yes	Yes	Yes	Not reported
Waarts et al 2012	Yes	Yes	Yes	Not reported	Yes	Yes	Yes	Yes	Yes	Not reported
Waarts et al 2014	Yes	Yes	Yes	Yes	Yes	Not clear	Yes	Yes	Yes	Not reported
Waarts et al 2016	Yes	Yes	Yes	Not reported	Yes	Yes	Yes	Yes	Yes	Not reported
Walsh 2004	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported	Yes	Not reported
Wilson 2010	Yes	Not reported	Yes	Not clear	Yes	Not reported	Yes	Not clear	Yes	Not reported

Annex D: Coding tools

Coding tool for review question 1[99]

Effect Size Coding Tool	ID	Description
General	CODER ID	Initials of coder
	STUDY ID	Unique study (dataset) identification #
	REPORT ID	Unique report (publication) identification #
	ES NO	Effect size number (to distinguish multiple ES per study)
	AUTHOR	First Author
	PUB DATE	Publication date
	PUB TYPE	Publication type
	PUB O	Specify other publication type
	IND	Independent evaluation
	IND COM	
	COM G	General comments
Study Context	COUNTRY	Country in which research took place
	REGION	World Bank Region a country belongs to
	LOC	Research location with country of study
	COMMOD	Certified commodity
	PROD SYSTEM	Production system
	COM SC	Study context comments
Intervention Basics (Short Version)	CS	Certification scheme or standard
	POP	Type of population under certification
	POP OTHER	Other Participant type
	POP ORG	Participants' organisation
	PRICE I	Price and contract interventions. Floor price. Pre-payment and credit. Assured purchases. **(Only certification related)**

[99] In addition we also extracted the information necessary for the calculation of effect sizes, such as sample sizes, standard deviations, test statistics, and regression coefficients.

	PRICE II	Specify interventions. Provide exact location in the text
	PREMIUM I	Premium (additional sum of money paid on top of the minimum price). Social premium for community interventions. Support to producers' organisations for use of social premium. **(Only certification related)**
	PREMIUM II	Specify interventions. Provide exact location in the text
	MARKET I	Market access interventions (Access to more lucrative market niches via label) **(Only certification related)**
	MARKET II	Specify interventions. Provide exact location in the text
	GAP I	Product quality interventions. Professional farm management and better farming practices/good agricultural practices. (including environmental standards) and associated technical assistance (capacity building) **(Only certification related)**
	GAP II	Specify interventions. Provide exact location in the text
	MANAGEMENT I	Producer group management. Monitoring of producer organisation practices and technical assistance (capacity building) to producer organisations and individual agricultural producers. **(Only certification related)**
	MANAGEMENT II	Specify interventions. Provide exact location in the text
	LABOUR I	Labour standards. Monitoring safe working conditions. Worker association training. Workers' rights. Monitoring and enforcing living/higher wages **(Only certification related)**
	LABOUR II	Specify interventions. Provide exact location in the text
	OTHER INTERV	Other Intervention type. **Non-certification related**
	COM OI	Other intervention comments. **Non-certification related**
	EXT ASSISTANCE	External assistance received by the treatment group

	COM IB	Intervention Basics comments.
Study design	S TYPE	Design type
	OTHER TYPE	Other study type
	S METH	Study Methods
	S METH OT	Study methods other
	S UNIT	Sampling unit
	SAM SIZE	Sample size
	COM SD	Study design comments
Risk of Bias	SELECTION BIAS	Mechanism of assignment
	GROUP EQIV	Group equivalence
	MOTIVATION BIAS	Hawthorne and John Henry effects - being observed changes behaviour
	SELECTIVE OUTCOMES	Selective reporting of outcomes
	SELECTIVE ANALYSIS	Selective reporting of analysis
	OTHER BIAS	Other sources of bias
	UOA	Unit of analysis
	OVERALL ROB ASSESSMENT	Final risk of bias score

Coding tool for review question 2

Qualitative Coding Tool	ID	Description
General	CODER ID	Initials of coder
	STUDY ID	Unique study (dataset) identification #
	REPORT NO	Unique report (publication) number
	REPORT ID	Unique report (publication) identification #
	AUTHOR	Author (s)
	PUB DATE	Publication date
	PUB TYPE	Publication type
	PUB O	Specify other publication type
	FUNDER	Funding agency
	IND	Independent evaluation
	COM G	General comments
Study Context	COUNTRY	Country in which research took place
	REGION	World Bank Region in which country belongs
	LOC	Research location with country of study
	COMMOD	Certified commodity
	COM SC	Study context comments
Intervention Characteristics	CS	Certification scheme or standard
	POP	Type of population under certification
	POP OTHER	Other Participant type
	POP ORG	Participants' organisation
	THEORY	Any description/statement of program theory explaining how certification is expected to work on the ground in order to achieve changes in the lives of the targeted population
	PRICE I	Price and contract interventions. Floor price. Pre-payment and credit. Assured purchases. **(Only certification related)**
	PRICE II	Specify interventions. Provide exact location in the text
	PREMIUM I	Premium (additional sum of money paid on top of the minimum price). Social premium for community interventions. Support to producers' organisations for use of social premium. **(Only certification related)**

	PREMIUM II	Specify interventions. Provide exact location in the text
	MARKET I	Market access interventions (Access to more lucrative market niches via label) **(Only certification related)**
	MARKET II	Specify interventions. Provide exact location in the text
	GAP I	Product quality interventions. Professional farm management and better farming practices/good agricultural practices. (including environmental standards) and associated technical assistance (capacity building) **(Only certification related)**
	GAP II	Specify interventions. Provide exact location in the text
	MANAGEMENT I	Producer group management. Monitoring of producer organisation practices and technical assistance (capacity building) to producer organisations and individual agricultural producers. **(Only certification related)**
	MANAGEMENT II	Specify interventions. Provide exact location in the text
	LABOUR I	Labour standards. Monitoring safe working conditions. Worker association training. Workers' rights. Monitoring and enforcing living/higher wages **(Only certification related)**
	LABOUR II	Specify interventions. Provide exact location in the text
	OTHER INTERV	Other Intervention type - **Non-certification related**
	COM OI	Other intervention comments. **Non-certification related**
	COM IB	Intervention Basics comments.
Study Methods	METH	Main study methods
	METH O	Other study methods
Data Extraction		
Implementation Dynamics	IMPL	Presence of material related to the implementation of the standards (trainings & new practices; participant targeting and selection; services; monitoring & auditing)
	TRAIN	Material related to training and new practices (i.e. good agricultural practices)
	SEL	Material related to participants' targeting and selection (self-selection of producers included, i.e. if and how producers self-select themselves in or out of the certification programme)
	SERV	Material related to certification related services (i.e. distribution of inputs, such as chemicals and fertilisers; credit services; etc)
	MON	Material related to monitoring and auditing
	COSTS	Material related to implementation costs of certification programmes
	F PREM	Material related to financial premium use (i.e. amount of money paid to producers on top of the market price)
	S PREM	Material related to social premium use

	IMPL O	Other material relevant to implementation dynamics
	COM IM	Implementation dynamics comments
Ditributional Dynamics	DIS	Presence of material on distributional dynamics
	BEN	Material related to distribution of benefits & investments
	ELIT	Material related to elite capture (i.e. when resources transferred designated for the benefit of the larger population are usurped by a few individuals of superior status–be it economic, political, educational, ethnic, or otherwise)
	ADOPT E	Material related to even adoption of standards (Was adoption even across the population? i.e. between men and women? between different social groups?)
	DIS O	Other material relevant to distributional dynamics
	COM DIS	Distribution dynamics comments
Causal mechanisms and Barriers and Facilitators	B&F I	Presence of material on causal mechanisms and barriers and facilitators
	COOP	Material related to cooperative management/performance
	PART	Material related to participants' reception/motivation
	ADOPT	Material related to effective/non-effective adoption (were producers able to adopt, apply and comply with certification standards?)
	DEM S	Material related to demand for certification services
	GOV	Material related to labour legislation, government interventions and regulations, and incentives to invest in working conditions improvements
	INF	Material related to social infrastructure (existence or lack of cooperative movement; community services; healthcare and education services)
	MARKET	Material related to price differentiation, premiums and new markets, and markets' capacity to remunerate
	VOL	Material related to market volatility
	DEM P	Material related to demand for certified products (i.e lack of demand or excessive offer of certified products)
	CERT COST	Material related to production cost for certified products (i.e labour, input, organic production)
	B&F O	Other material relevant to causal mechanisms and barriers and facilitators
	COM B&F	Causal mechanisms and barriers and facilitators comments
Summary for synthesis	COM DE	General comments on data extraction
	SUM	Summary of main findings, lessons, points of study that can be useful for synthesis
Risk of Bias	R QUEST I	Clarity of research questions
	R QUEST II	(Summary and location in text)
	R APP I	Clear justification of research approach (i.e. ethnography, grounded theory, mixed methods, etc)
	R APP II	(Summary and location in text)
	APP&Q	Assessment of whether the approach is appropriate for the research question

R CON I	Clear description of study context
R CON II	(Summary and location in text)
REFLEX I	Clear description of the researcher(s)' role
REFLEX II	(Summary and location in text)
SAM I	Clear description of research participants' sampling methods
SAM II	(Summary and location in text)
SAM&Q	Assessment of whether the sampling strategy is appropriate for the research question
SIT I	Clear justification of research site(s) selection
SIT II	(Summary and location in text)
SIT & Q	Assessment of whether the research site selection is appropriate for the research question
DATA COL I	Clear description of data collection methods
DATA COL II	(Summary and location in text)
DATA &Q	Assessment of whether the data collection is appropriate for the research question
ANALYSIS I	Clear description of analysis method
ANALYSIS II	(Summary and location in text)
ANALYSIS&Q	Assessment of whether the data analysis is appropriate for the research question
EVIDENCE I	Assessment of whether the claims made are sufficiently supported by evidence
EVIDENCE II	(Summary and location in text)
TRIAN I	Data triangulation (cross verification of data from more than two sources)
TRIAN II	(Summary and location in text)
COM ROB	Risk of bias assessment comments

Annex E: Example of extracts excluded and included for synthesis

Example of excluded extract: "Funds from the Fairtrade Premium had also been used to help combat HIV/AIDS, including co-finance for HIV testing and awareness-raising." (Smith, 2010:49)

The above extract on the use of Fairtrade Premium in Ghanaian plantations was excluded, on the basis that it does not provide enough substantive relevant evidence on implementation or distributional dynamics, nor on contextual factors that can affect the use of the premium by workers.

Example of included extract: "Fairtrade had had another important impact on the social and legal status of Haitians in the Dominican Republic. Funds from the Premium were being used to process passports and working visas, giving them protection from the regular mass expulsions of migrant workers by Dominican authorities. It also reduced the cost of travelling to and from Haiti, as they no longer had to pay “coyotes” to get them across the border illegally. However, obtaining the right to stay and work in the Dominican Republic did not give automatic rights to workers’ children, as Dominican law required children to have a Dominican birth certificate in order to attend school and access other public services. As a result some Haitians were resorting to paying Dominicans to “adopt” their children so they could obtain a birth certificate and gain full citizenship." (Smith, 2010: 50)

This extract on the use of Fairtrade Premium in plantations in the Dominican Republic, on the other hand, was included and coded under the codes "Social Premium" and "Public Institutions", as we consider that it provides enough substantial relevant evidence on all our key themes of interest for the following reasons: it describes specifically not only how the Fairtrade premium was used (process passports and working visas) , but also provides a link with direct effects (provide protection; reduction of travelling costs), and therefore it is was considered relevant for implementation dynamics. Further, it provides insights on how a specific group of workers (illegal immigrants from Haiti) benefited from a certification input and therefore it was relevant for distributional dynamics. Finally, it provides contextual information (expulsions of migrant workers by Dominican authorities) which was deemed important to understand contextual barriers and facilitators that can affect CS effectiveness.

Annex F: Unit of analysis-adjusted meta-analysis results

As explained in Section 3 in the main body of the report, we have conducted all meta-analyses both with and without adjusting for unit of analysis errors. This annex presents the results of the meta-analyses after adjusting for unit of analysis errors. As laid out in the main body of the report, not all studies reported the necessary information and we thus had to rely on assumptions in a number of cases. In particular, fifteen studies did not report all of the necessary information. We therefore made assumptions about the relevant number of clusters and/or cluster sizes in order to make the necessary adjustments for these studies. As a robustness check we also re-ran the analyses excluding all of the studies for which we had to impute values, and there was no substantive difference between the two sets of analyses. As the analyses using imputed values are more complete, only those results are presented here.

Comparison of unadjusted and UoA-adjusted effect sizes

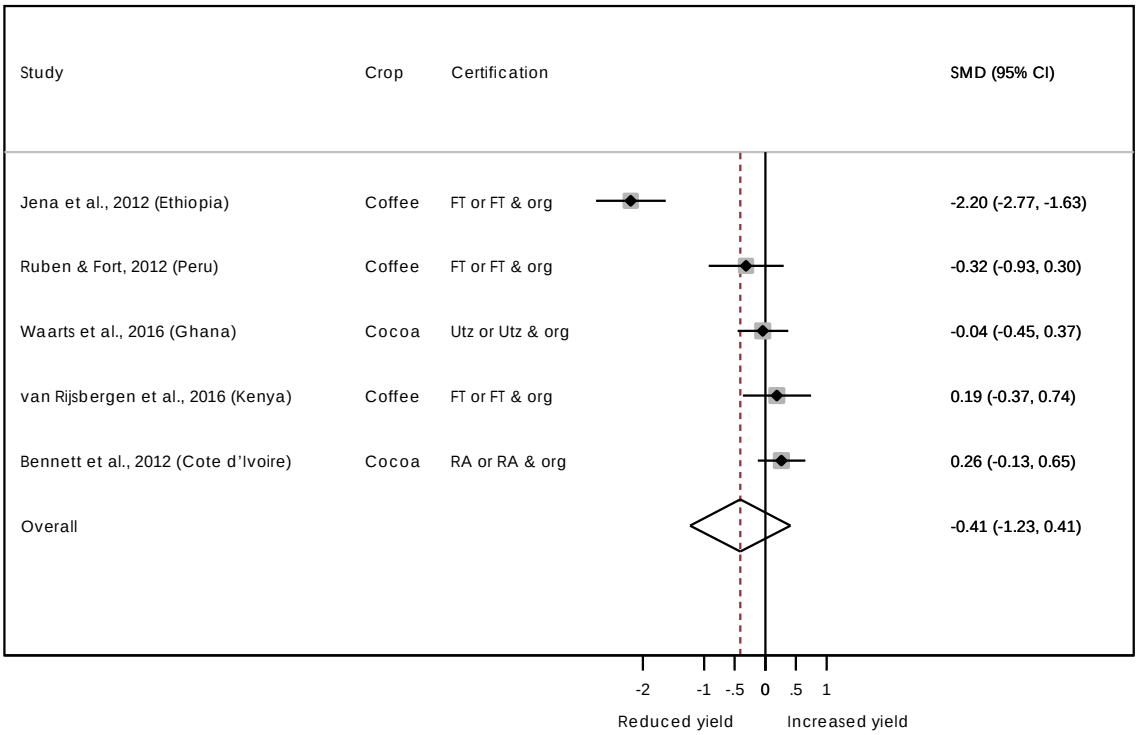
Outcome	Unadjusted values			Adjusted values		
	SMD	95%-CI LB	95%-CI UB	SMD	95%-CI LB	95%-CI UB
Yield	-0.42	-1.23	0.39	-0.41	-1.23	0.41
Price	0.28	0.08	0.47	0.32	0.15	0.49
Income cert. prod.	0.22	0.03	0.41	0.26	0.08	0.44
Wages	-0.26	-0.46	-0.06	-0.23	-0.42	-0.03
Total HH income	0.13	-0.06	0.32	0.16	-0.05	0.36
Assets/wealth	0.05	-0.15	0.26	-0.03	-0.31	0.36
Illness	-0.15	-0.32	0.03	-0.15	-0.40	0.10
Schooling	0.12	0.00	0.24	0.14	-0.05	0.32

Generally, adjusting for unit of analysis errors increases the standard errors of effect size estimates, producing wider confidence intervals. In other words, the effect size estimates become less precise. In some cases the differences in standard errors for individual effect size estimates is large. However, adjusting the standard errors made no substantive difference to the pooled effect size estimates for all outcomes apart from schooling. By substantive difference we mean that the adjustment did not affect the direction of effect or its statistical significance, nor did it produce large changes in the size of the estimated effect. The table below shows the SMD, as well as the lower and upper bounds of their 95%-confidence intervals. In the case of schooling, though, what had previously been a statistically significant positive pooled effect is no

longer statistically significant once the unit of analysis adjust has been made, though the estimated effect remains positive. Another partial exception is the estimated pooled effect size for wealth, which does change sign – though the pooled effect remains statistically non-significant.

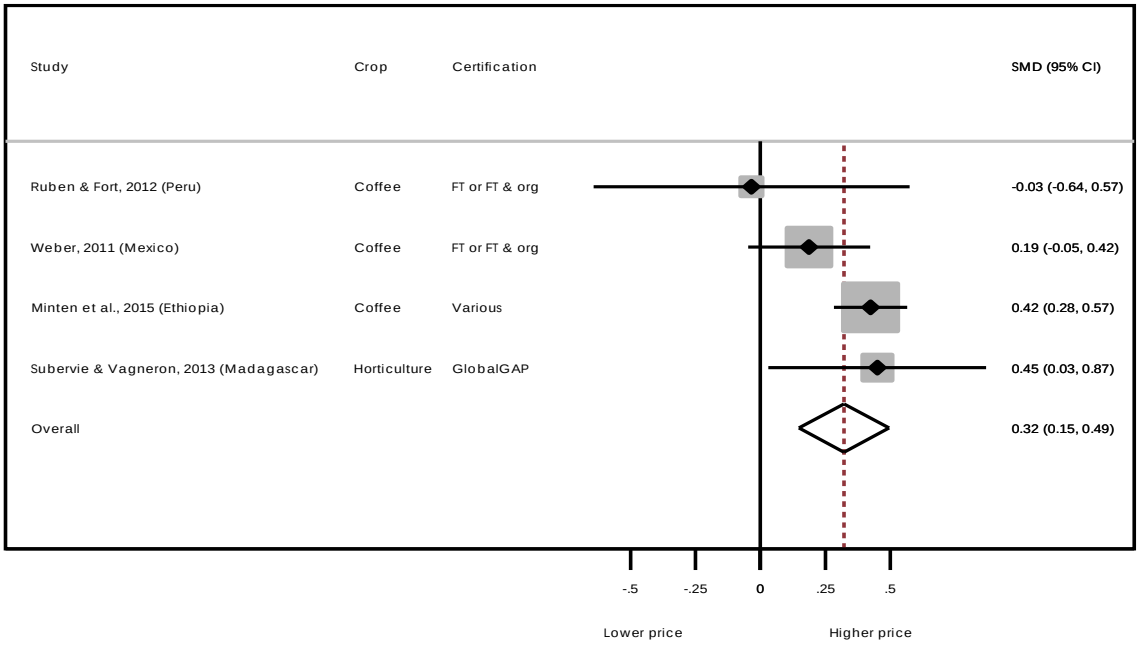
Below we present the forest plots displaying meta-analysis results after the standard errors of effect size estimates have been corrected for unit of analysis errors. As before, diagnostic statistics are given below the plots. Please also note that for presentational reasons the scales differ from plot to plot.

UoA-adjusted forest plot for yields



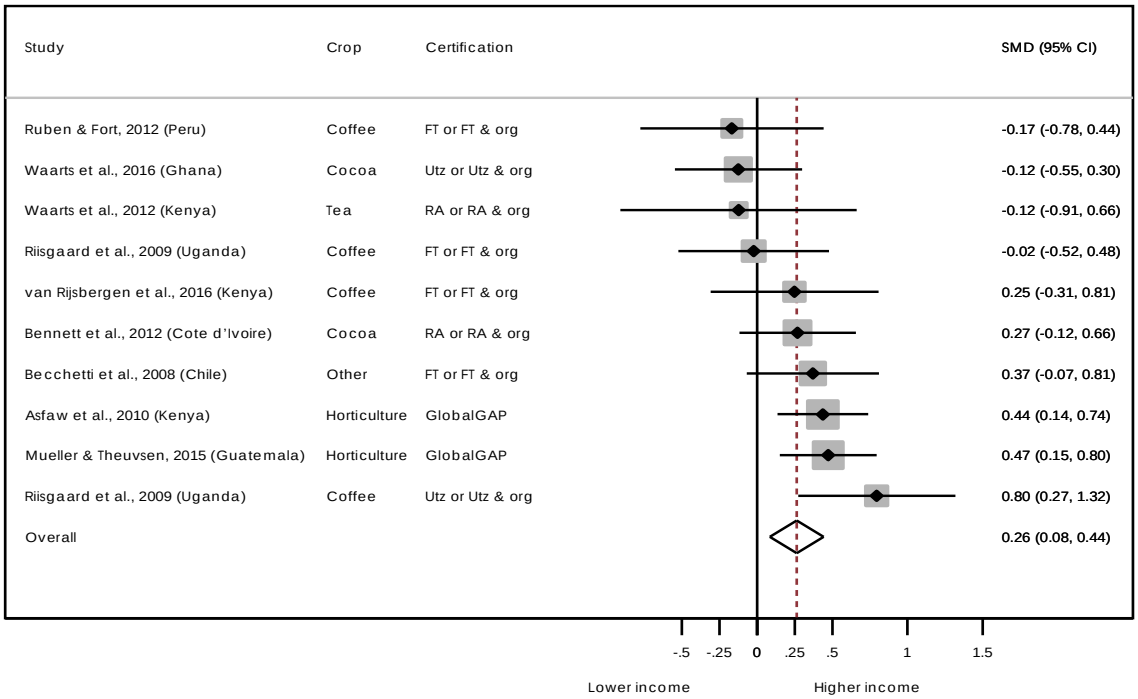
Diagnostics: Heterogeneity chi-squared = 55.30 (d.f. = 4) p = 0.000; I-squared (variation in ES attributable to heterogeneity) = 92.8%; Estimate of between-study variance Tau-squared = 0.8031; Test of ES=0: z= 0.98, p = 0.326.

UoA-adjusted forest plot for price



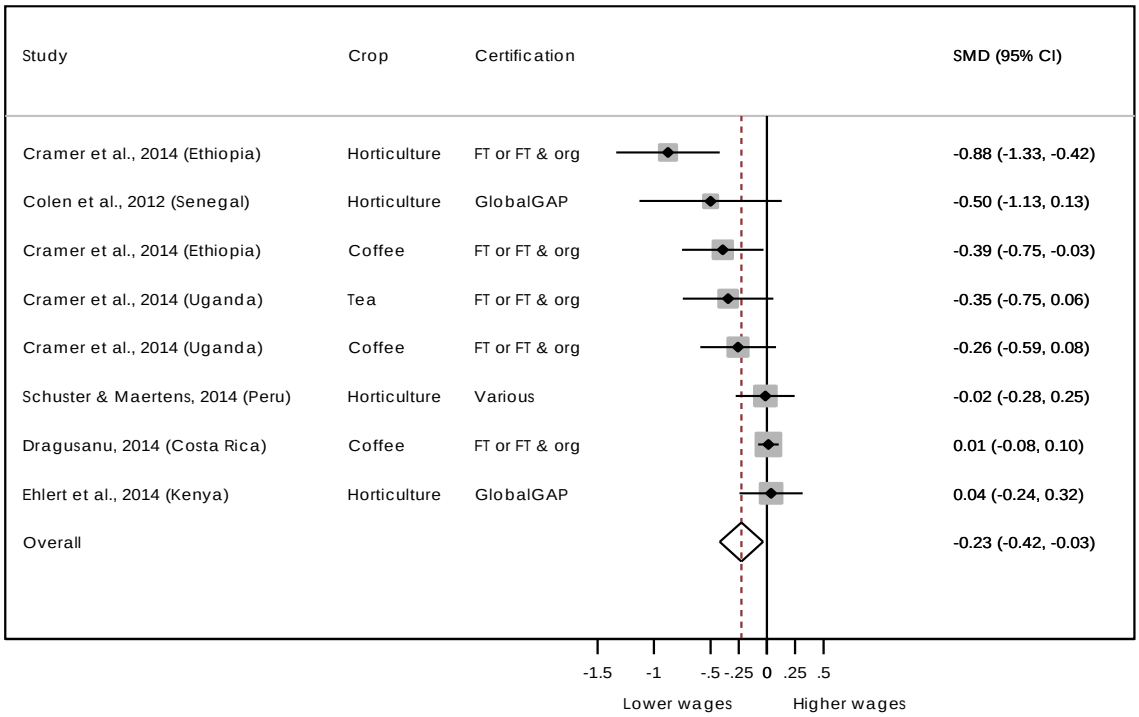
Diagnostics: Heterogeneity chi-squared = 4.65 (d.f. = 3) p = 0.200; I-squared (variation in ES attributable to heterogeneity) = 35.4%; Estimate of between-study variance Tau-squared = 0.0110; Test of ES=0: z= 3.64, p = 0.000.

UoA-adjusted forest plot for income from certified production



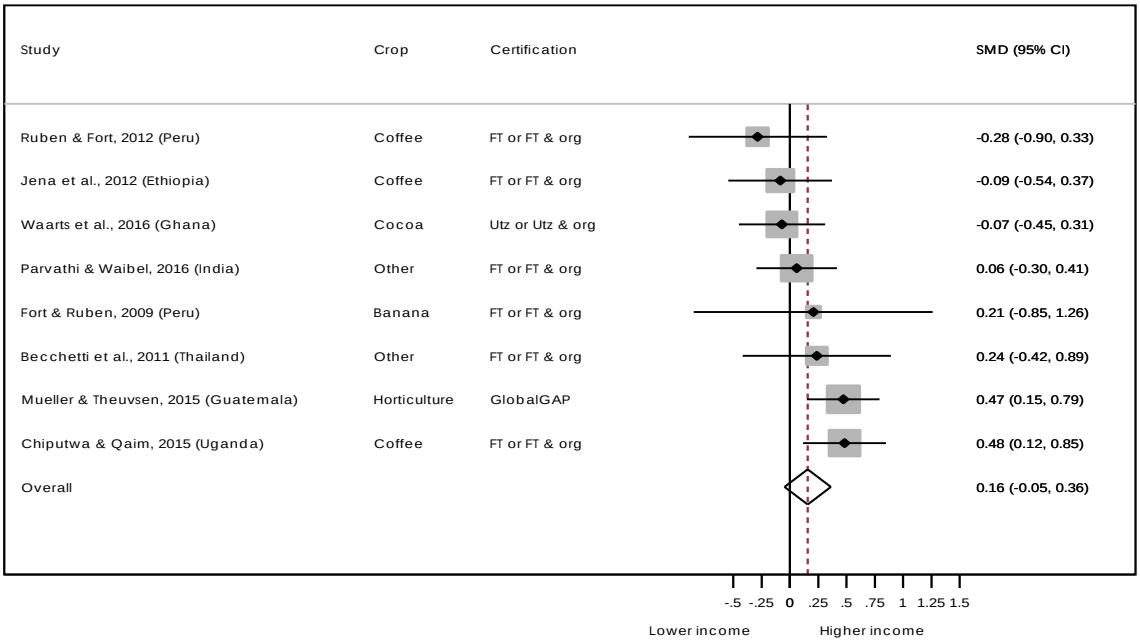
Diagnostics: Heterogeneity chi-squared = 14.32 (d.f. = 9) p = 0.111; I-squared (variation in ES attributable to heterogeneity) = 37.2%; Estimate of between-study variance Tau-squared = 0.0299; Test of ES=0: z= 2.87, p = 0.004.

UoA-adjusted forest plot for wages



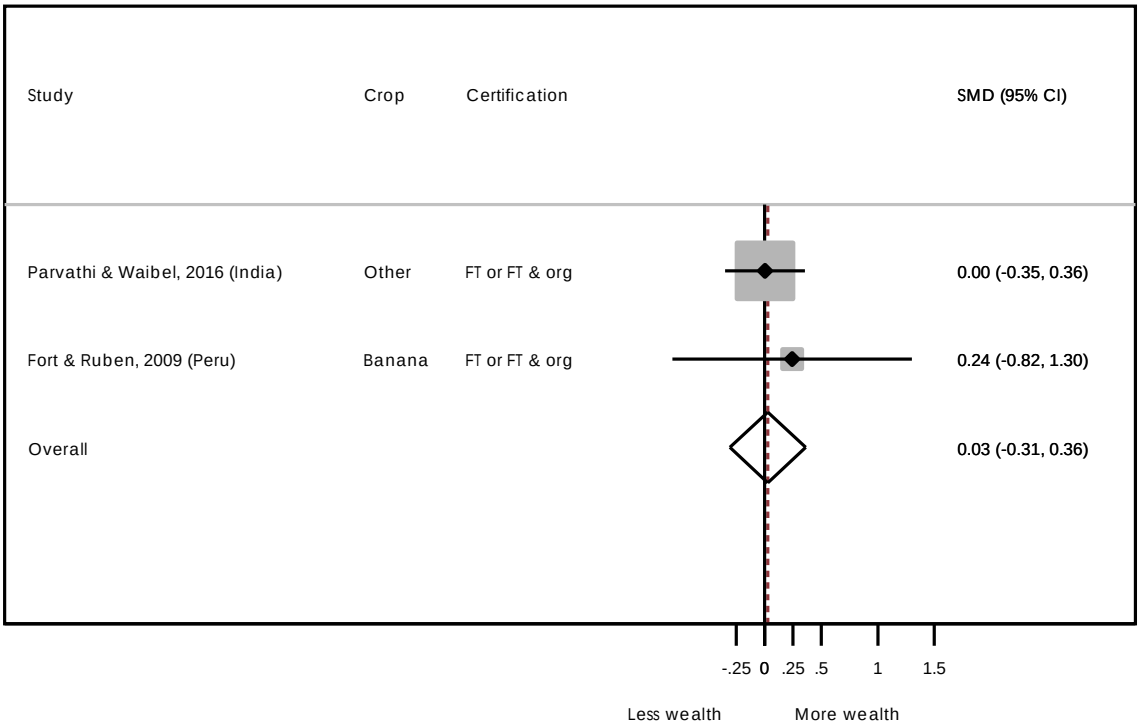
Diagnostics: Heterogeneity chi-squared = 23.56 (d.f. = 7) p = 0.001; I-squared (variation in ES attributable to heterogeneity) = 70.3%; Estimate of between-study variance Tau-squared = 0.0464; Test of ES=0: z= 2.31, p = 0.021.

UoA-adjusted forest plot for total household income (net and gross)



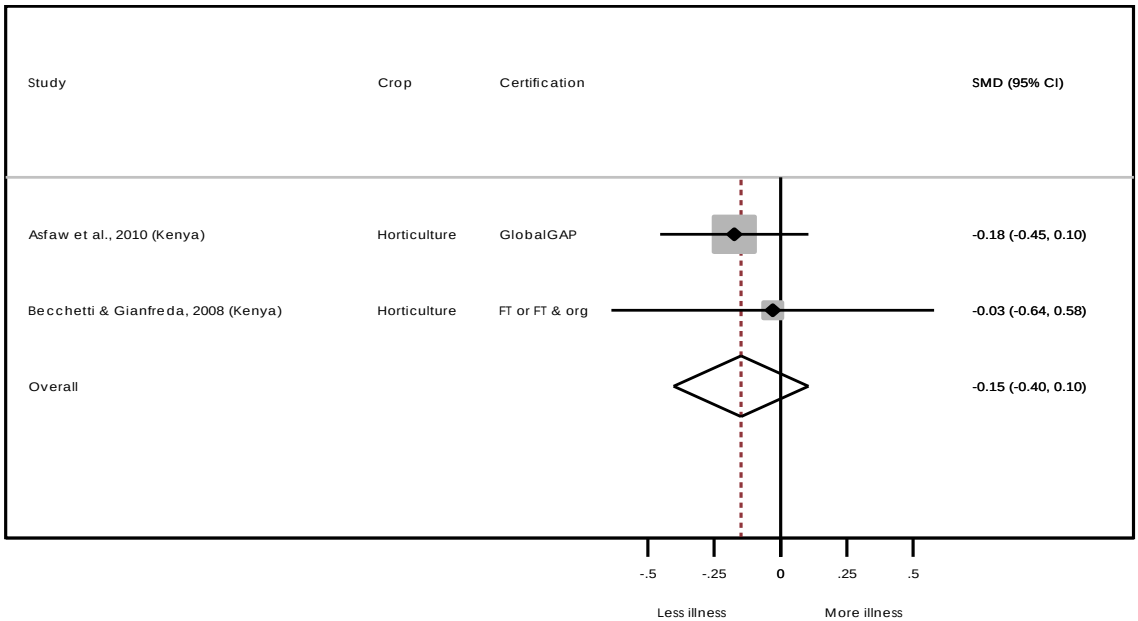
Diagnostics: Heterogeneity chi-squared = 11.39 (d.f. = 7) p = 0.122; I-squared (variation in ES attributable to heterogeneity) = 38.5%; Estimate of between-study variance Tau-squared = 0.0320; Test of ES=0: z= 1.49, p = 0.136.

UoA-adjusted forest plot for wealth



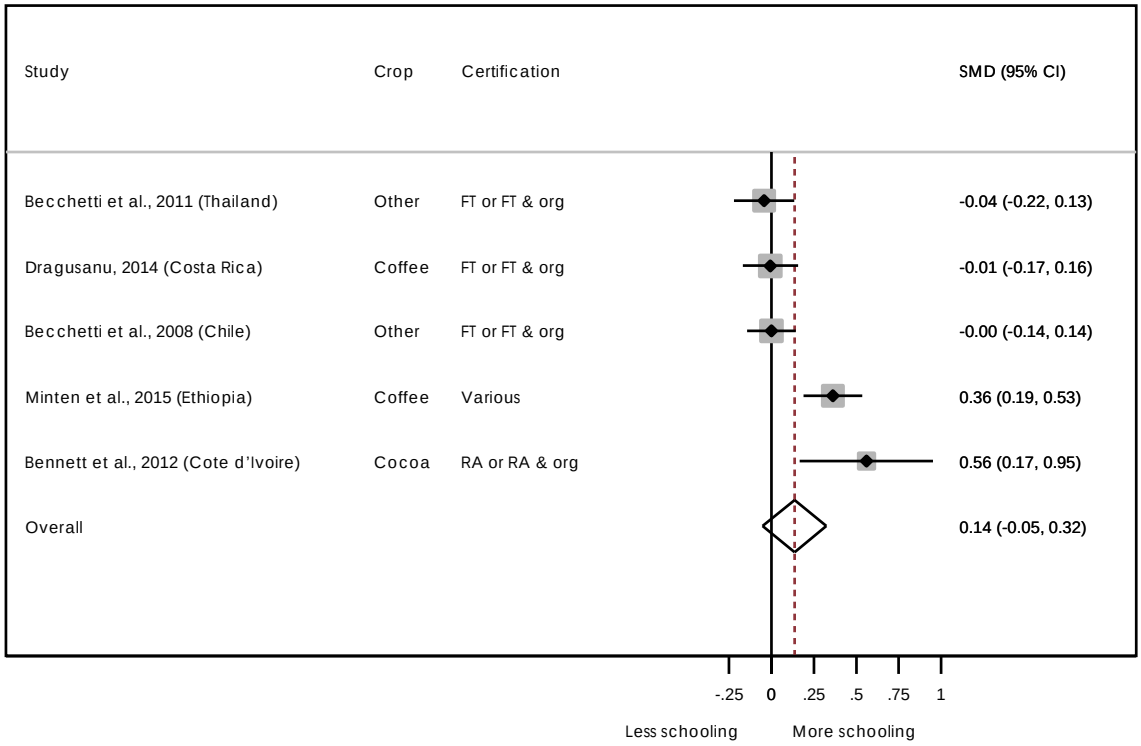
Diagnostics: Heterogeneity chi-squared = 0.18 (d.f. = 1) p = 0.673; I-squared (variation in ES attributable to heterogeneity) = 0.0%; Estimate of between-study variance Tau-squared = 0.0000; Test of ES=0: z= 0.16, p = 0.876.

UoA-adjusted forest plot for illness



Diagnostics: Heterogeneity chi-squared = 0.18 (d.f. = 1) p = 0.671; I-squared (variation in ES attributable to heterogeneity) = 0.0%; Estimate of between-study variance Tau-squared = 0.0000; Test of ES=0: z= 1.16, p = 0.248.

UoA-adjusted forest plot for schooling



Diagnostics: Heterogeneity chi-squared = 19.89 (d.f. = 4) p = 0.001; I-squared (variation in ES attributable to heterogeneity) = 79.9%; Estimate of between-study variance Tau-squared = 0.0345; Test of ES=0: z= 1.42, p = 0.156.

Annex G: Contribution of each study to the qualitative synthesis

The table below demonstrates the overall contribution of each included study to the three evidence themes. Each evidence theme was supported by evidence from at least more than 50% of the included studies. More studies provided evidence on the evidence themes of implementation dynamics and causal mechanisms and barriers and facilitators than did for distributional dynamics.

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Abarca-Orozco 2015	Yes	No	Yes
Aidenvironment 2016	Yes	No	Yes
Amekawa 2001	No	No	Yes
Arce 2009	Yes	Yes	Yes
Asfaw et al 2009	Yes	No	Yes
Babin 2012	No	No	Yes
Bacon 2005	Yes	Yes	Yes
Bagama et al 2014	Yes	No	Yes
Bakker 2014	Yes	Yes	Yes
Balineau 2011	Yes	No	Yes
Barham and Weber 2012 (Mexico)	Yes	No	No
Barham and Weber 2012 (Peru)	Yes	No	No

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Beall 2012	Yes	No	Yes
Bergeron 2010	Yes	Yes	Yes
Besky 2014	Yes	Yes	Yes
Beuchelt 2009	Yes	No	Yes
Bonanno and Cavalcanit 2012	Yes	No	Yes
Brown 2012	Yes	Yes	Yes
Carimentrand and Ballet2010	No	Yes	No
Cepeda et al 2013	Yes	Yes	Yes
CESU 2012 (Ecuador)	Yes	Yes	No
CESU 2012 (Kenya)	Yes	Yes	Yes
CESU 2012 (Ghana)	Yes	Yes	No
CESU 2012 (Peru; Coffee)	Yes	Yes	Yes
CESU 2012 (Peru; Banana)	Yes	Yes	Yes
CESU 2012 (India)	Yes	Yes	No
Chiputwa et al 2015	Yes	No	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Cofre et al 2012	Yes	Yes	Not clear
Cramer et al 2014a (Ethiopia; Coffee)	No	Yes	Yes
Cramer et al 2014a (Ethiopia; Flowers)	No	Yes	No
Cramer et al 2014a (Uganda; Coffee)	No	Yes	Yes
Cramer et al 2014a (Uganda; Tea)	No	Yes	No
Cramer et al 2014b (Uganda)	Yes	Yes	Yes
Cramer et al 2014b (Ethiopia)	Yes	Yes	Yes
Dannenberg and Nduru 2013	Yes	No	Yes
Dolan 2010	No	No	No
Donovan and Poole 2014a	Yes	Yes	Yes
Donovan and Poole 2014b	Yes	No	Yes
Dowdall 2012	Yes	Yes	Yes
Dragusanu and Nunn 2014	Yes	Yes	No
Ellery 2010	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Fairtrade 2013 (Tea)	Yes	Yes	Yes
Fairtrade 2013 (Groundnuts)	Yes	Yes	Yes
Fairtrade 2013 (Sugar)	Yes	Yes	Yes
Fairtrade 2015 (Bananas)	Yes	Yes	Yes
Fairtrade 2015 (Cotton)	Yes	Yes	Yes
Fairtrade 2015 (Tea)	Yes	Yes	Yes
Fayet and Vermeulen 2014	Yes	No	Yes
Franccesconi and Ruben 2014	Yes	Yes	Yes
Fraser et al 2014	Yes	Yes	Yes
Garcia et al 2014	Yes	No	Yes
Getz and Schreck 2006	Yes	Yes	Yes
Gómez-Cardona 2012	Yes	No	Yes
Hanson et al 2012	No	Yes	Yes
Heller 2010	Yes	Yes	Yes
Herman 2010	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Jaffee 2006	Yes	No	Yes
Jari et al 2013	Yes	Yes	Yes
Jena et al 2012	Yes	No	Yes
Kariuki 2014	No	No	Yes
Köhne 2014	No	No	No
Laroche et al 2012	Yes	Yes	Yes
Larsen et al 2014	Yes	No	No
Loconto and Simbua 2010	Yes	Yes	Yes
Luetchford 2008	No	No	Yes
Lyall 2014	Yes	No	Yes
Lyon 2005	Yes	No	Yes
Lyon 2006	Yes	Yes	No
Lyon 2007	Yes	No	No
Lyon et al 2010	Yes	Yes	Yes
Makita 2011	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Makita 2012	Yes	No	Yes
Melkeraaen 2009	Yes	Yes	Yes
Mendez 2002	Yes	Yes	Yes
Milford 2004	Yes	Yes	Yes
Milford 2014	Yes	No	Yes
Minten et al 2015	Yes	No	Yes
Moberg 2008a	Yes	Yes	Yes
Moberg 2008b	Yes	Yes	Yes
Moore 2010	Yes	No	Yes
Moyo and Mugabe 2014	Yes	No	Yes
Naylor 2014	Yes	No	Yes
Nelson et al 2002	Yes	Yes	Yes
Nelson and Martin 2013a	Yes	Yes	Yes
Nelson and Martin 2013 b (Kenya; Wage workers)	Yes	Yes	Yes
Nelson and Martin 2013b (Kenya; Producers)	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Nelson and Martin 2013b (India)	Yes	Yes	Yes
Nelson and Martin 2013b (Ecuador)	Yes	Yes	No
Nelson and Martin 2013b (Ghana)	Yes	Yes	No
Ouma 2010	Yes	No	Yes
Pollack 2006	No	Yes	Yes
Pongratz 2007	Yes	Yes	Yes
Quaedvlieg et al 2014	Yes	No	Yes
Raynolds 2012	Yes	Yes	Yes
Raynolds 2014	Yes	No	Yes
Riisgaard et al 2009 (Uganda, Coffee)	Yes	Yes	Yes
Riisgaard et al 2009 (Uganda, Tea)	Yes	Yes	Yes
Riisgaard et al 2009 (Kenya, Tea)	Yes	Yes	Yes
Riisgaard et al 2009 (Kenya; Coffee)	Yes	Yes	Yes
Riisgaard et al 2009 (Ethiopia)	Yes	Yes	Yes
van Rijn et al 2016 (Ghana)	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
van Rijn et al 2016 (Colombia)	Yes	Yes	Yes
van Rijn et al 2016 (Dominican Republic)	Yes	Yes	Yes
Romanoff 2010	Yes	No	No
Ronchi 2002	Yes	Yes	Yes
Rotter 2011	Yes	Yes	Yes
Roy and Thorat 2008	Yes	No	Yes
Rueda and Lambin 2013	Yes	Yes	Yes
Ryan 2011	Yes	Yes	Yes
Said-Allsopp and Tallontire 2014	Yes	Yes	Yes
Schelly 2011	Yes	Yes	Yes
Schoonhoven-Speijer 2012	Yes	Yes	Yes
Schuster and Maertens 2013	No	No	Yes
Sen 2009	Yes	Yes	Yes
Setrini 2011	Yes	Yes	Yes
Schreck 2002	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Silva-Castaneda 2012	Yes	No	Yes
Smith et al 2004	No	Yes	Yes
Smith 2007	Yes	No	Yes
Smith 2010 (Ghana)	Yes	Yes	Yes
Smith 2010 (Ecuador)	Yes	Yes	Yes
Smith 2010 (Various)	Yes	Yes	Yes
Smith 2010 (Windward Islands)	Yes	Yes	Yes
Smith 2010 (Dominican Republic)	Yes	Yes	Yes
Staib 2012	Yes	Yes	Yes
Staricco and Ponte 2015	Yes	Yes	Yes
Stathers and Gathuthi 2013	Yes	Yes	Yes
Stenn 2015	Yes	Yes	Yes
Subervie and Vagneron 2013	Yes	Yes	No
Sutton 2014	Yes	Yes	Yes
Terstappen 2010	Yes	Yes	Yes

Study	Evidence on implementation dynamics (n=127)	Evidence on distributional dynamics (n=97)	Evidence on causal mechanisms and barriers and facilitators (n=123)
Trauger 2014	Yes	Yes	Yes
TWIN 2012 (Côte d'Ivoire)	Not clear	No	Not clear
TWIN 2012 (Kenya)	Yes	No	Yes
TWIN 2012 (Malawi)	Yes	No	Yes
TWIN 2013	No	Yes	Yes
Valkila and Nygren 2008	Yes	Yes	Yes
Valkila 2009	Yes	Yes	Yes
Vásquez-León 2010	Yes	No	Yes
Waarts et al 2012	Yes	No	Yes
Waarts et al 2014	Yes	Yes	Yes
Waarts et al 2016	Yes	Yes	No
Walsh 2004	Yes	Yes	Yes
Wilson 2010	Yes	Yes	Yes

References

References to Studies Included for RQ1 Only

Anteneh et al (2014)

Anteneh, A., Muradian, R., and Ruben, R. (2014). The more the better? The effects of multiple certification on smallholder coffee livelihood: evidence from southern Ethiopia. In: Hoebink, P., Ruben, R., Elbers, W., and van Rijsbergen, B., ed. 2014, *The impact of coffee certification on smallholder farmers in Kenya, Uganda and Ethiopia.* Radboud University Nijmegen: Centre for International Development Issues Nijmegen.

Asfaw et al (2010)

Asfaw, S., Mithoefer, D., and Waibel, H. (2008). EU private agrifood standards in African high-value crops: pesticide use and farm-level productivity. *12th Congress of the European Association of Agricultural Economists (EAAE)*, pp.26-29.

Asfaw, S., Mithoefer, D., and Waibel, H. (2008). EU food safety standard, pesticide use, and farm-level productivity: the case of high-value crops in Kenya. *Journal of Agricultural Economics*, 16(3), pp.645-667.

Asfaw, S., Mithoefer, D., and Waibel, H. (2009a) Investment compliance with GlobalGAP standards: does it pay off for small-scale producers in Kenya? *Quarterly Journal of International Agriculture*, 48(4), pp.337-362.

Asfaw, S., Mithoefer, D., and Waibel, H. (2009b). Food safety standards and farmers' health: evidence from Kenya's export vegetable growers. *International Association of Agricultural Economists*, Beijing, August 16-22.

Asfaw, S., Mithoefer, D., and Waibel, H. (2010). Agrifood supply chain, private-sector standards, and farmers' health: evidence from Kenya. *Agricultural Economics*, 41, pp.251-263.

Becchetti and Gianfreda (2008)

Becchetti, L., and Constantino, M. (2008). The effects of Fair Trade on affiliated producers: an impact analysis on Kenyan farmers. *World Development*, 36(5), pp. 823-842.

Becchetti, L., and Gianfreda, G. (2008). Does Consumption Heal Producers? The Effect of Fair Trade on Marginalised Producers' Health in Kenya. *Cahiers Agricultures*, 19, pp.23-27.

Becchetti et al (2008)

Becchetti, L., and Castriota, S (2008). Is Fairtrade honey sweeter? An empirical analysis of the effect of affiliation on productivity. *ECINE Working Paper Series*, WP2008-104, Society for the Study of Economic Inequality.

Becchetti, L., Castriota, S., and Michetti, M. (2008). *Child labour and consumer responsibility: an impact study*. Working Paper 103, Society for the Study of Economic Inequality.

Becchetti et al (2011)

Becchetti, L., Conzo, P, and Gianfreda, G. (2008). Market access, organic farming and productivity: the determinants of creation of economic value on a sample of Fair Trade affiliated Thai farmers. *EconomEtica Working Paper*, 5, CIPEERSI.

Becchetti, L., Conzo, P, and Gianfreda, G. (2011). Market access, organic farming and productivity: the determinants of creation of economic value on a sample of Fair Trade affiliated Thai farmers. *The Australian Journal of Agricultural and Resource Economics*, 56, pp.117–140.

Becchetti, L., Conzo, P., and Pisani, F. (2011). Virtuous interactions in removing exclusion: The link between foreign market access and access to education. *Journal of Development Studies*, 47(9), pp.1431-1451.

Bennett et al (2012)

Bennett, M., Francesconi, G.N., Giovannucci, D., and Daitchman, J. (2012). *Cote d'Ivoire Cocoa: A COSA survey of Rainforest Alliance certified farms*. Committee on Sustainability Assessment.

Colen et al (2012)

Colen, L., Maertens, M., and Swinnen, J. (2012). Private standards, trade and poverty: GlobalGAP and horticultural employment in Senegal. *World Economy,* 35(8), pp.1073-1088.

Ehlert et al (2014)

Ehlert, C.R., Mithoefer, D., and Waibel, H. (2014). Worker Welfare on Kenyan Export Vegetable Farms. *Food Policy*, 46, pp.66-73.

Fort and Ruben (2009)

Fort, R., and Ruben, R. (2009). The impact of Fair Trade on banana producers in northern Peru. Paper prepared for: *International Association of Agricultural Economists Conference*, Beijing, China, August 16-22, pp.42-55.

Kamau et al (2010)

Kamau, M.W., Mose, L., Fort, R., and Ruben, R. (2010). T*he impact of certification on smallholder coffee farmers in Kenya: the case of the Utz certification program*. Paper prepared for: The Joint 3rd African Association of Agricultural Economists (AAAE) and 48th Agricultural Economists Association of South Africa (AEASA) Conference, Cape Town, South Africa, September 19-23.

Kuit et al (2016) Uganda and Vietnam [2 studies]

Kuit, M., Guinée, L., Van Anh. P. (2016). *Estimating the impact of implementation of the 4C entry level standard in Uganda and Vietnam*. Kuit Consultancy.

Mueller and Theuvsen (2015)

Mueller, A.K., and Theuvsen, L. (2015). *Financial literacy and food safety standards in Guatemala: The heterogeneous impact of GlobalGAP on farm income*. Global Food Discussion Papers, 63. Germany: Georg-August-University of Göttingen.

Parvathi and Waibel (2016)

Parvathi, P., and Waibel, H. (2016). Organic Agriculture and Fair Trade: A Happy Marriage? Evidence from Black Pepper Farmers in India. *World Development*, 77, pp.206–220.

Ruben and Zuniga-Arias (2011)

Ruben, R., and Zuniga, G. (2011). How standards compete: comparative impact of coffee certification schemes in Northern Nicaragua. *Supply Chain Management: An International Journal*, 16(2), pp.98-109.

Ruben and Fort (2012)

Ruben, R., and Fort, R. (2012). The impact of fair trade certification for coffee farmers in Peru. *World Development*, 40(3), pp.570-582.

Ruben et al (2014)

Ruben, R., Hoebink, P., Ngutu, M.H., Mutwiwa, U., Njuguna, S., Elbers, W., van Kempen, L., and van Rijsbergen, B. (2014). The Impact of Coffee Certification on Smallholder Farmers in Kenya. In: Hoebink, P., Ruben, R., Elbers, W., and van Rijsbergen, B., ed. 2014, *The impact of coffee certification on smallholder farmers in Kenya, Uganda and Ethiopia.* Radboud University Nijmegen: Centre for International Development Issues Nijmegen.

Schuster and Maertens (2014)

Schuster, M., and Maertens, M. (2014). *Do labor standards benefit workers in horticultural export chains in Peru?* Bioeconomics Working Paper Series, (2014/10), Belgium: University of Leuven.

Van Rijsbergen et al. (2016)

van Rijsbergen, B., Elbers, W., Ruben, R., and Njuguna, S.N. (2016). The ambivalent impact of coffee certification on farmers' welfare: a matched panel approach for cooperatives in central Kenya. *World Development*, 77, pp.277-292.

Weber (2011)

Weber, J.G. (2011). How Much More Do Growers Receive for Fair Trade-Organic Coffee? *Food Policy*, 36(5), pp.678-685.

References to Studies Included for both RQ1 and RQ2

Balineau (2011)

Balineau, G. (2011). Le renforcement des organisations de producteurs de coton au Mali: Enjeux, impact et leçons du commerce équitable pour la privatisation de la filière. *Canadian Journal of Development Studies*, 32(3), pp.297-312.

Barham and Weber (2012)

Barham, B.L., and Weber, J.G. (2012). The Economic Sustainability of Certified Coffee: Recent Evidence from Mexico and Peru. *World Development*, 40(6), pp.1269-1279.

Cepeda et al (2013)

Cepeda, D., Pound, B., Nelson, V., Kajman, G., Cabascango, D., Martin, A., Chile, M., Posthumus, H., Caza, G., Mejia, I., Montenegro, F., Ruup, L., Velastegui, G.A., Tiaguaro, Y., Valverde, M., and Ojeda, A. (2013). *Assessing the Poverty Impact of Sustainability Standards: Ecuadorian Cocoa*. University of Greenwich: Natural Resources Institute.

Chiputwa and Qaim (2014)

Chiputwa, B., and Qaim, M. (2014). Sustainability standards, gender, and nutrition among smallholder farmers in Uganda. *GlobalFood Discussion Papers*, RTG1666, Georg-August-University of Göttingen.

Cramer et al (2014) [4 studies]

Cramer, C., Johnston, D., Oya, C., and Sender, J. (2014). *Fairtrade Employment and Poverty Reduction in Ethiopia and Uganda*. SOAS: The University of London.

Dragusanu and Nunn (2014)

Dragusanu, R., and Nunn, N. (2014). *The impacts of Fair Trade certification: Evidence from coffee producers in Costa Rica*. Harvard University.

García et al (2014)

García, C., García, J., Ochoa, G., Mora, J., and Castellanos, J. (2014). *Impact Evaluation of UTZ Certified Coffee Program in Colombia. (2008-2012)*. Colombia: CRECE.

Jena et al (2012)

Jena, P.R., Chichaibelu, B.B., Stellmacher, T., and Grote, U. (2012). The impact of coffee certification on small-scale producers' livelihoods: a case study from the Jimma Zone, Ethiopia. *Agricultural Economics*, 43(4), pp. 429-440.

Minten et al (2015)

Minten, B., Dereje, M., Engeda, E., and Tamru, S. (2015). *Who benefits from the rapidly increasing Voluntary Sustainability Standards? Evidence from Fairtrade and Organic certified coffee in Ethiopia.* IFPRI Working Paper 71.

Nelson et al (2013)

Nelson, V., Opoku, K., Martin, A., Bugri, J., and Posthumus, H. (2013). *Assessing the poverty impact of sustainability standards: Fairtrade in Ghanaian cocoa*. University of Greenwich: Natural Resources Institute.

Riisgaard et al (2009) [2 studies]

Riisgaard, L., Michuki, G., Gibbon, P., and Bolwig, S. (2009). *The performance of voluntary standard schemes from the perspective of small producers in East-Africa*. Copenhagen, Denmark: Danish Institute for International Studies.

Roy and Thorat (2008)

Roy, D, and Thorat, A. (2008). Success in High Value Horticultural Export Markets for the Small Farmers: The Case of Mahagrapes in India. *World Development*, 36(10), pp.1874-1890.

Van Rijn et al (2016) [3 studies]

van Rijn, F.C., Judge, L.O., Fort, R., Koster, T., Waarts, Y.R., and Ruben, R. (2016). *Fairtrade certification in the hired banana sector*. LEI report 2015-056, LEI Wageningen UR.

Stathers and Gathuthi (2013)

Stathers, T., and Gathuthi, C. (2013). *Poverty impact of social and environmental voluntary standard systems in Kenyan tea*. University of Greenwich: Natural Resources Institute.

Subervie and Vagneron (2013)

Subervie, J., and Vagneron, I. (2013). A Drop of Water in the Indian Ocean? The Impact of GlobalGap Certification on Lychee Farmers in Madagascar. *World Development*, 50, pp. 57-73.

Waarts et al (2012)

Waarts, Y., Ge, L., Ton, G., and Jansen, D. (2012). *Sustainable tea production in Kenya: impact assessment of Rainforest Alliance and Farmer Field School training*. LEI report 2012-043, LEI Wageningen UR.

Waarts et al (2016)

Waarts, Y., Ingram, V., Liderhof, V., Puister-Jansen, L., and van Rijn, F, and Aryeetey, R. (2016). *Impact of UTZ certification on cocoa producers in Ghana, 2011 to 2014*. LEI Wageningen UR.

References to Studies Included for RQ2 only

Abarca-Orozco, S.J. (2015). *Production and marketing innovations in Fair Trade and organic coffee cooperatives in the Cordoba-Huatusco corridor in Veracruz, Mexico*. PhD. Iowa State University.

Aidenvironment. (2016). *Evaluation of Utz in the Indonesian cocoa sector*. Amsterdam: Aidenvironment.

Amekawa, Y. (2001). Can a public GAP approach ensure safety and fairness? A comparative study of Q-GAP in Thailand. *Journal of Peasant Studies*, 40(1), pp.189-217.

Arce, A. (2009). Living in Times of Solidarity: Fair Trade and the Fractured Life Worlds of Guatemalan Coffee Farmers. *Journal of International Development*, 21, pp.1031-1041.

Asfaw, S., Mithofer, D., and Waibel, H. (2009). Investment in Compliance with GlobalGAP Standards: Does It Pay Off for Small-Scale Producers in Kenya? *Quarterly Journal of International Agriculture*, 48(4), pp.337-362.

Babin, N. (2012). *Agrarian Change, Agroecological Transformation and the Coffee Crisis in Costa Rica*. PhD. University of California.

Bacon, C.M. (2005). *Confronting the coffee crisis: Nicaraguan farmers use of cooperative, Fair Trade and agroecological networks to negotiate livelihoods and sustainability*. PhD. University of California, Santa Cruz.

Bagama, F., Hoebink, P., Elber, W., van Kempen, L., and van Rijsbergen, B. (2014). The Impact of UTZ Certification on Smallholder Farmers in Uganda. In: Hoebink, P., Ruben, R., Elbers, W., and van Rijsbergen, B., ed. 2014, *The impact of coffee certification on smallholder farmers in Kenya, Uganda and Ethiopia.* Radboud University Nijmegen: Centre for International Development Issues Nijmegen.

Bakker, P. (2014). *Do Smallholders Benefit from Group Certification? A Case Study on the Social and Economic Effects of Rainforest Alliance Group Certification on Small-Scale Tea Farmers in Misiones, Argentina*. Masters. Wageningen UR.

Beall, E. (2012). Smallholders in Global Bioenergy Value Chains and Certification: Evidence from Three Case Studies. *Environment and Natural Resources Management Working Paper, 50.*

Bergeron, M. (2010). "*Para dar mas titulo a las mujeres": Le potentiel d'empowerment du commerce equitable de cafe pour les productrices costaricaines*. Masters. University of Ottawa.

Besky, S. (2014). *The Darjeeling distinction: labor and justice on fair-trade tea plantations in India*. Berkeley: University of California Press.

Beuchelt, T., Zeller, M., and Oberthur, T. (2009). Justified hopes or utopian thinking? The suitability of coffee certification schemes as a business model for small-scale producers. In: *International Association of Agricultural Economists Conference*, Beijing, China, 16-22 August 2009.

Bonanno, A., and Cavalcanti, J.S.B. (2012). Globalization, food quality and labor: the case of grape production in North-eastern Brazil. *International Journal of Sociology of Agriculture and Food*, 19(1), pp.37-55.

Brown, S. (2012). One hundred years of labor control: violence, militancy, and the Fairtrade banana commodity chain in Colombia. *Environment and Planning A*, 45(11), pp.2572-2591.

Carimentrand, A., and Ballet, J. (2010). When Fair Trade increases unfairness: The case of quinoa from Bolivia. *Cahier FREE, 5.*

Cofre, G., Riquelme, I., Engler, A., and Jara-Rojas, R. (2012). Adopción de Buenas Prácticas Agrícolas (BPA): costo de cumplimiento y beneficios percebidos entre productores de fruta fresca. *IDESIA*, 30(3), pp.37-45.

Cramer, C., Johnston, D., Oya, C., and Sender, J. (2014). Fairtrade cooperatives in Ethiopia and Uganda: uncensored. *Review Of African Political Economy*, 41, pp.S115-S127.

Center for Evaluation of the Saarland University (CESU). (2012). *Assessing the Impact of Fairtrade on Poverty Reduction through Rural Development*. Saarland University.

Dannenberg, P., and Nduru, G.M. (2013). Practices in international value chains: the case of the Kenyan fruit and vegetable chain beyond the exclusion debate. *Tijdschrift voor economische en sociale geografie*, 104(1), pp.41-56.

Dolan, C.S. (2010). Virtual moralities: the mainstreaming of Fairtrade in Kenyan tea fields. *Geoforum*, 41(1), pp.33-43.

Donovan, J., and Poole, N. (2014). Changing asset endowments and smallholder participation in higher value markets: Evidence from certified coffee producers in Nicaragua. *Food Policy*, 44, pp.1-13.

Donovan, J., and Poole, N. (2014). Partnerships in Fairtrade coffee: a close-up look at how buyers and NGOs build supply capacity in Nicaragua. *Food Chain*, 4(1), pp.34-48.

Dowdall, C.M. (2012). *Small farmer market knowledge and specialty coffee commodity chains in Western Highlands Guatemala*. PhD. Florida International University.

Dragusanu, R., and Nunn, N. (2014). *The Impacts of Fair Trade Certification: Evidence From Coffee Producers in Costa Rica*. Harvard Papers.

Ellery, R. (2010). *Spilling the beans: Exploring women's empowerment in fair trade coffee production*. Masters. University of Guelph.

Fairtrade Foundation. (2013). *Branching Out: Fairtrade in Malawi - Monitoring the impact of Fairtrade on five certified organisations*. London: Fairtrade Foundation.

Fairtrade Foundation. (2015). *Equal Harvest Removing the Barriers to Women's Participation in Smallholder Agriculture*. London: Fairtrade Foundation.

Fayet, L., and Vermeulen, W.J.V. (2014). Supporting smallholders to access sustainable supply chains: lessons from the Indian cotton supply chain. *Sustainable Development*, 22(5), pp.289-310.

Francesconi, G.N., and Ruben, R. (2014). FairTrade's theory of change: an evaluation based on the cooperative life cycle framework and mixed methods. *Journal Of Development Effectiveness*, 6(3), pp.268-283.

Fraser, J., Fisher, E., and Arce, A. (2014). Reframing 'Crisis' in Fair Trade Coffee Production: Trajectories of Agrarian Change in Nicaragua. *Journal of Agrarian Change*, 14(1), pp.52-73.

Getz, C., and Schreck, A. (2006). What organic and Fair Trade labels do not tell us: towards a place-based understanding of certification. *International Journal of Consumer Studies,* 30(5), pp. 490-501.

Gómez-Cardona, S. (2012). Tensions of Organic Markets for Colombian Coffee Growers: The Case of Valle del Cauca. *Cuadernos de Desarrollo Rural*, 9(68), pp.65-85.

Hanson, L., Terstappen, V., Bacon, C.M., Leung, J., Ganem-Cuenca, A., Flores, S.R.D., and Rojas, M.A.M. (2012). Gender, health, and Fairtrade: insights from a research-action programme in Nicaragua. *Development in Practice*, 22(2), pp.164-179.

Heller, A. (2010). *Why organizations matter: Certification experiences of coffee producer groups in Guatemala*. PhD. Colorado State University.

Herman, A. (2010). Connecting the complex lived worlds of Fairtrade. *Journal of Environmental Policy and Planning*, 12(4), pp.405-422.

Jaffee, D.S. (2006). *Producing justice: Fair trade coffee, livelihoods and the environment*. PhD. The University of Wisconsin - Madison.

Jari, B., Snowball, J.D., and Fraser, G.C.G. (2013). Is Fairtrade in commercial farms justifiable? Its impact on commercial and small-scale producers in South Africa. *Agrekon*, 52(4), pp. 66-88.

Kariuki, I.M. (2014). Transition to certification schemes and implications for market access: GlobalGAP perspectives in Kenya. *Agricultural Sciences*, 5(12), pp.1100-1111.

Köhne, M. (2014). Multi-stakeholder initiative governance as assemblage: Roundtable on Sustainable Palm Oil as a political resource in land conflicts related to oil palm plantations. *Agriculture and Human Values*, 31(3), pp.469-480.

Laroche, K., Jimenez, R., and Nelson, V. (2012). *Assessing the Impact of Fairtrade for Peruvian Cocoa Farmers*. University of Greenwich: Natural Resources Institute.

Larsen, R.K., Jiwan, N., Rompas, A., Jenito, J., Osbeck, M., and Tarigan, A. (2014). Towards 'hybrid accountability' in EU biofuels policy? Community grievances and competing water claims in the Central Kalimantan oil palm sector. *Geoforum*, 54, pp.295-305.

Loconto, A.M., and Simbua, E.F. (2012). Making Room for Smallholder Cooperatives in Tanzanian Tea Production: Can Fairtrade Do That? *Journal of Business Ethics*, 108(4), pp.451-465.

Luetchford, P. (2008). *Fair trade and a global commodity: coffee in Costa Rica*. London: Pluto Press.

Lyall, A. (2014). *Assessing the Impacts of Fairtrade on Worker-Defined Forms of Empowerment on Ecuadorian Flower Plantations*. Switzerland: Fairtrade International and Max Havelaar Foundation.

Lyon, S., Bezaury, J.A., and Mutersbaugh, T. (2010). Gender equity in Fairtrade-organic coffee producer organizations: cases from Mesoamerica. *Geoforum*, 41(1), pp.93-103.

Lyon, S. (2006). Migratory imaginations: the commodification and contradictions of shade grown coffee. *Social Anthropology*, 14(3), pp.377-390.

Lyon, S. (2007). Fair Trade Coffee and Human Rights in Guatemala. *Journal of Consumer Policy*, 30(3), pp.241-261.

Lyon, S. (2005). *Maya coffee farmers and the fair trade commodity chain*. PhD. Emory University.

Makita, R. (2011). A confluence of Fair Trade and organic agriculture in southern India. *Development in Practice*, 21(2), pp.205-217.

Makita, R. (2012). Fair Trade Certification: The Case of Tea Plantation Workers in India. *Development Policy Review*, 30(1), pp.87-107.

Melkeraaen, S. (2009). *Black economic empowerment in the South African wine industry: fair trade, power relations and socio-economic rights*. Masters. The Norwegian University Library of Life Sciences.

Mendez, V.E. (2002). *Fair Trade Networks in Two Coffee Cooperatives of Western El Salvador.* PhD. Colorado State University.

Milford, A. (2004). *Coffee, co-operatives and competition: the impact of fair trade*. CMI Report R 2004 6. Bergen: Christian Michelsen Institute.

Milford, A.B. (2014). Co-operative or coyote? Producers' choice between intermediary purchasers and Fairtrade and organic co-operatives in Chiapas. *Agriculture and Human Values,* 31(4), pp. 577-591.

Moberg, M. (2005). Fair trade and Eastern Caribbean banana farmers: rhetoric and reality in the anti-globalization movement. *Human Organization*, 64(1), pp.4-15.

Moberg, M. (2008). *Slipping away: banana politics and fair trade in the Eastern Caribbean*. New York: Berghahn Books.

Moore, L.B. (2010). *Reading Tea Leaves: The Impact of Mainstreaming Fair Trade*. Working Paper 10-106. London School of Economics.

Moyo, P., and Mugabe, T.C. (2014). An appraisal of the socio-economic impacts and challenges of fair trade among emerging farmers in the Eastern Cape, South Africa. *Journal of Human Ecology*, 47(2), pp.201-207.

Naylor, L.B. (2014). *Decolonial autonomies: Fair trade, subsistence and the everyday practice of food sovereignty in the highlands of Chiapas*. PhD. University of Oregon.

Nelson, V., and Martin, A. (2013). *Final Technical Report: Assessing the Poverty Impact of Sustainability Standards*. University of Greenwich: Natural Resources Institute.

Ouma, S. (2010). Global Standards, Local Realities: Private Agrifood Governance and the Restructuring of the Kenyan Horticulture Industry. *Economic Geography*, 86(2), pp.197-222.

Pollack, N. (2006). *Women's empowerment in Fair Trade coffee co-operatives in Oaxaca, Mexico*. Masters. Saint Mary's University.

Pongratz-Chander, K. (2007). *Fair trade coffee in Costa Rica: A new model for sustainable development?* PhD. Northern Arizona University.

Quaedvlieg, J., Roca, M.G., and Ros-Tonen, M.A.F. (2014). Is Amazon nut certification a solution for increased smallholder empowerment in Peruvian Amazonia? *Journal of Rural Studies*, 33, pp.41-55.

Raynolds, L.T. (2012). Fair Trade Flowers: Global Certification, Environmental Sustainability, and Labor Standards. *Rural Sociology*, 77(4), pp.493-519.

Raynolds, L.T. (2014). Fairtrade, certification, and labor: global and local tensions in improving conditions for agricultural workers. *Agriculture and Human Values*, 31(3), pp.499-511.

Romanoff, S. (2010). Shade coffee in biological corridors: potential results at the landscape level in El Salvador. *Culture and Agriculture*, 32(1), pp.27-41.

Ronchi, L. (2002). *The Impact of Fair Trade on Producers and Their Organisations: A Case Study with Coocafe in Costa Rica*. PRUS Working Paper, 11. University of Sussex.

Rotter, S. (2000). *A just harvest: A naturalistic evaluation of a fair trade coffee project in San Lucas Toliman, Guatemala*. Masters. University of Guelph.

Rueda, X., and Lambin, E.F. (2013). Responding to Globalization: Impacts of Certification on Colombian Small-Scale Coffee Growers. *Ecology and Society*, 18(3), 21.

Ryan, O. (2011). *Chocolate nations: living and dying for cocoa in West Africa*. London: Zed Books.

Said-Allsopp, M., and Tallontire, A. (2014). Enhancing Fairtrade for women workers on plantations: insights from Kenyan agriculture. *Food Chain*, 4(1), pp.66-77.

Schelly, E. (2011). *Representation and partnership: A case study of the worker committee on fair trade certified farms in the Ecuadorian cutflower industry*. Masters. Colorado State University.

Schoonhoven-Speijer, M. (2012). *Maintaining a sustainable livelihood; An analysis of the effects of UTZ certification on market access, risk reduction and livelihood strategies of Kenyan coffee farmers*. Masters. Radboud University Nijmwegen.

Schuster, M., and Maertens, M. (2013). Do Private Standards Create Exclusive Supply Chains? New Evidence from the Peruvian Asparagus Export Sector. *Food Policy*, 43, pp.291-305.

Sen, D. (2009). *From illegal to organic: Fair trade-organic tea production and women's political futures in Darjeeling, India.* PhD. Rutgers, The State University of New Jersey.

Setrini, G. (2011). *Global niche markets and local development: clientelism and Fairtrade farmer organizations in Paraguay's sugar industry*. PhD. Massachusetts Institute of Technology.

Shreck, A. (2002). *Just bananas? A Fair Trade alternative for small-scale producers in the Dominican Republic*. PhD. Colorado State University.

Silva-Castaneda, L. (2012). A forest of evidence: third-party certification and multiple forms of proof - a case study of oil palm plantations in Indonesia. *Agriculture And Human Values*, 29(3), pp.361-370.

Smith, J. (2007). The Search for Sustainable Markets: The Promise and Failures of Fair Trade. *Culture & Agriculture*, 29(2), pp.89-99.

Smith, S., Auret, D., Barrientos, S., Dolan, C., Kleinbooi, K., Njobvu, C., Opondo, M., and Tallontire, A. (2004). *Ethical trade in African horticulture: gender, rights and participation*. University of Sussex: Institute of Development Studies.

Smith, S. (2010). Fairtrade Bananas: A Global Assessment of Impact. University of Sussex: Institute of Development Studies.

Staib, P.W. (2012). *Coffee and the countryside: Small farmers and sustainable development in Las Segovias de Nicaragua*. PhD. The University of New Mexico.

Staricco, J.I., and Ponte, S. (2015). Quality regimes in agro-food industries: a regulation theory reading of Fair Trade wine in Argentina. *Journal of Rural Studies*, (38), pp.65-76.

Stenn, T.L. (2015). *The Cultural and Political Intersection of Fair Trade and Justice: Managing a Global Industry*. New York: Palgrave Macmillan.

Sutton, S. (2014). *Voice, Choice and Governance: The Case of Tanzania's Fairtrade Co-operatives*. PhD. Queen Mary, University of London.

Terstappen, V.L. (2010). *A case study of gender, health, and Fair Trade in Nicaragua*. Masters. University of Saskatchewan.

Trauger, A. (2014). Is Bigger Better? The Small Farm Imaginary and Fair Trade Banana Production in the Dominican Republic. *Annals of the Association of American Geographers*, 104(5), pp.1082-1100.

TWIN. (2013). *Empowering Women Farmers in Agricultural Value Chains*. London: TWIN.

TWIN. (2012). *Making International Supply Chains Work for Smallholder Farmers: A comparative study of six fair trade value chains*. London:Fairtrade Foundation.

Valkila, J., and Nygren, A. (2008). Impacts of Fair Trade-certification on Coffee Farmers, Cooperatives and Laborers, in Nicaragua. *Agriculture and Human Values*, 27(3), pp.321-333.

Valkila, J. (2009). Fair Trade organic coffee production in Nicaragua: Sustainable development or a poverty trap? *Ecological Economics*, 68, pp.3018-3025.

Vásquez-León, M. (2010). Free markets and fair trade, collective livelihood struggles, and the cooperative model: two case studies from Paraguay. *Latin American Perspectives*, 37(6), pp.53-73.

Waarts, Y., Lan, G., and Giel, T. (2014). *From training to practice: Midterm evaluation of the UTZ-Solidaridad smallholder tea programme in Malawi*. LEI report 2013-051. LEI Wageningen UR.

Walsh, J.M. (2004). *Fair Trade in the Fields: Outcomes for Peruvian Coffee Producers*. Masters. Massachusetts Institute Of Technology.

Wilson, B.R. (2010). Indebted to Fair Trade? Coffee and crisis in Nicaragua. *Geoforum*, 41(1), pp.84-92.

References to Excluded Studies

Addae, C. (2005). *The expansion of the concept of Fair Trade to increase the economic welfare of rural farmers in Ghana: A case study of the Kuapa Kokoo Farmers' Cooperative*. Masters. University of Massachusetts Lowell.

Aidoo, R., and Fromm, I. (2015). Willingness to adopt certifications and sustainable production methods among small-scale cocoa farmers in the Ashanti Region of Ghana. *Journal of Sustainable Development*, 8(1), pp.33-43.

Almeida Dias, R.A., Almeida Dias, R.E.B., Prado, A.S., Pereira, S.P., de Aguiar, C.M.G., and de Aguiar Neto, A.P. (2011). Perfil Do Trabalhador Rural Em Fazendas Cafeeiras Certificadas e Não Certificadas. In: VII Simpósio de Pesquisa dos Cafés do Brasil 22 a 25 de Agosto de 2011, Araxá – MG.

Alvarado, J. (2009). Fair trade in Mexico and abroad: an alternative to the Walmartopia? *Journal of Business Ethics*, 88(Supp. 2), pp.301-317.

Alves, S.A.O. (2011). *Sustentabilidade da agroindústria de palma no estado do Pará*. PhD. Universidade de São Paulo.

Arana-Coronado, J.J., Bijman, J., Omta, O., and Oude-Lansink, A. (2013). Contractual arrangements and food quality certifications in the Mexican avocado industry. *Spanish Journal of Agricultural Research*, 11(1), pp.3-18.

Arnould, E.J., Plastina, A., and Ball, D. (2006). Market Disintermediation and Producer Value Capture: The Case of Fair Trade Coffee in Nicaragua, Peru and Guatemala. In: *Product and Market Development for Subsistence Marketplaces: Consumption and Entrepreneurship Beyond Literacy and Resource Barriers*. Chicago: University of Illinois at Chicago.

Arnould, E.J., Plastina, A., and Ball, D. (2009). Does Fair Trade Deliver on Its Core Value Proposition? Effects on Income, Educational Attainment, and Health in Three Countries. *Journal of Public Policy & Marketing,* 28(2), pp.186-201.

Asfaw, S., Mithofer, D., and Waibel, H. (2009). EU Food Safety Standards, Pesticide Use and Farm-Level Productivity: The Case of High-Value Crops in Kenya. *Journal of Agricultural Economics*, 60(3), pp.645-667.

Asian Development Bank. (2012). *Case Studies on Cross-Boarder Ecotrade*. Philippines: Asian Development Bank.

Assogba, S.C.G., Tossou, R.C., Lebailly, P., and Magnon, Y. (2014). Sustainable intensification of agriculture in Benin: myth or reality? Lessons from organic and cotton made in Africa production systems. *International Journal of Agriculture Innovations and Research*, 2(5), pp.694-704.

Auerbach, N.N. (2012). Delicious Peace Coffee: Marketing Community in Uganda. *Review of Radical Political Economics*, 44(3), pp.337-357.

Babin, N. (2002). The Coffee Crisis, Fair Trade, and Agroecological Transformation: Impacts on Land-Use Change in Costa Rica. *Agroecology And Sustainable Food Systems*, 39(1), pp.99-129.

Bäckman, T. (2009). Fairtrade coffee and development: a field study in Ethiopia. *Minor Field Study Series no.188.* University of Lund: Department of Economics.

Bacon, C. (2005). Confronting the coffee crisis: can fair trade, organic, and specialty coffees reduce small-scale farmer vulnerability in Northern Nicaragua? *World Development*, 33(3), pp.497-511.

Bacon, C. (2012). Quality Revolutions, Solidarity Networks, and Sustainability Innovations: Following Fair Trade Coffee from Nicaragua to California. *Journal of Political Ecology*, 20, pp.70-179.

Bacon, C., Mendez, V., Flores, M., and Brown, M. (2008). Will "We" Achieve the Millennium Development Goals with Small-Scale Coffee Growers and Their Cooperatives? A Case Study Evaluating Fair Trade and Organic Coffee Networks in Northern Nicaragua. *The Center for Agroecology and Sustainable Food Systems Research Brief 12*. University of California, Santa Cruz.

Bacon, C.M., Mendez, V.E., Gomez, M.E.F., Stuart, D., and Flores, S.R.D. (2008). Are Sustainable Coffee Certifications Enough to Secure Farmer Livelihoods? The Millennium Development Goals and Nicaragua's Fair Trade Cooperatives. *Globalizations*, 5(2), pp.259-274.

Bacon, C.M., Sundstrom, W.A., Gomez, M.E.F., Mendez, V.E., Santos, R., Goldoftas, B., and Dougherty, I. (2014). Explaining the 'hungry farmer paradox': smallholders and fair trade cooperatives navigate seasonality and change in Nicaragua's corn and coffee markets. *Global Environmental Change*, (25), pp.133-149.

Bain, C. (2010). Governing the global value chain: GLOBALGAP and the Chilean fresh fruit industry. *International Journal of Sociology of Agriculture and Food*, 17(1), pp.1-23.

Bain, C., Ransom, E., and Worosz, M.R. (2010). Constructing credibility: using technoscience to legitimate strategies in agrifood governance. *Journal of Rural Social Sciences*, 25(3), pp.160-192.

Bain, C., Thiagarajan, D., and Busch, L. (2004). *The Relationship of Third-Party Certification to Sanitary/Phytosanitary (SPS) Measures and the International Agri-Food Trade Case Study: Ghana.* Maryland: Development Alternatives, Inc.

Baradaran, S., and Barclay, S. (2011). *Fair Trade and Child Labor*. *Columbia Human Rights Law Review,* (43)1, pp.1-63.

Barham, B.L., Callenes, M., Gitter, S., Lewis, J., and Weber, J. (2011). Fair Trade/Organic Coffee, Rural Livelihoods, and the "Agrarian Question": Southern Mexican Coffee Families in Transition. *World Development*, 39(1), pp.134-145.

Bassett, T.J. (2010). Slim pickings: Fairtrade cotton in West Africa. *Geoforum*, 41(1), pp.44-55.

Bastian, H. (2006). Keeping Fair Trade Fair in Mexico. *NACLA Report on the Americas*, 39(6), pp. 6-9.

Becchetti, L., and Costantino, M. (2008). The Effects of Fair Trade on Affiliated Producers: An Impact Analysis on Kenyan Farmers. *World Development*, 36(5), pp.823-842.

Becchetti, L., and Gianfreda, G. (0). Does Consumption Heal Producers? The Effect of Fair Trade on Marginalised Producers' Health in Kenya. *Cahiers Agricultures*, 19, pp.23-27.

Becchetti, L., and Michetti, M. (2008). *When Consumption Generates Social Capital: Creating Room for Manoeuvre for Pro-Poor Policies*. ECINEQ Working Paper 2008 – 88. Verona: Society for the Study of Economic Inequality.

Becchetti, L., and Michetti, M. (2010). When Fair Trade generates social capital by creating room for manoeuvre for pro-poor policies. *African Journal Of Business Management*, 4(14), pp.2903-2914.

Becchetti, L., Conzo, P., and Gianfreda, G. (2009). *Market Access, Organic Farming and Productivity: The Determinants of Creation of Economic Value on a Sample of Fair Trade Affiliated Thai Farmers*. CREI Working Paper no. 3/2009. Rome: Centro di Ricerca Interdipartimentale di Economia delle Istituzion,

Becchetti, L., Conzo, P., and Pisani, F. (2011). Virtuous interactions in removing exclusion: The link between foreign market access and access to education. *Journal of Development Studies*, 47(9), pp.1431-1451.

Berndt, C. (2007). *Is Fair Trade in Coffee Production Fair and Useful? Evidence from Costa Rica and Guatemala and Implications for Policy.* Mercatus Policy Series, Policy Comment 11. George Mason University.

Berntsen, E.T. (2010). *Fair trade in South Africa: an assessment of fair trade in the South African wine sector*. MSc. University of KwaZulu-Natal, Durban.

Beuchelt, T., Zeller, M., and Oberthur, T. (2009). Justified hopes or utopian thinking? The suitability of coffee certification schemes as a business model for small-scale producers. In: *International Association of Agricultural Economists Conference*, Beijing, China, pp.265-289.

Beuchelt, T.D., and Zeller, M. (2011). Profits and poverty: Certification's troubled link for Nicaragua's organic and fairtrade coffee producers. *Ecological Economics*, 70(7), pp.1316-1324..

Beuchelt, T.D., and Zeller, M. (2013). The role of cooperative business models for the success of smallholder coffee certification in Nicaragua: A comparison of conventional, organic and Organic-Fairtrade certified cooperatives. *Renewable Agriculture And Food Systems*, 28(3), pp.195-211.

Bezençon, V. (2011). Producers and the Fair Trade Distribution Systems: What Are the Benefits and Problems? *Sustainable Development*, 19(1), pp.60-70.

Bitzer, V., Glasbergan, P., and Arts, B. (2013). Exploring the potential of intersectoral partnerships to improve the position of farmers in global agrifood chains: findings from the coffee sector in Peru. *Agriculture and Human Values,* 30(1), pp.5-20.

Blowfield, M., and Gallet, S. (2001). *Ethical trade and sustainable rural livelihoods - case studies. Volta River Estates Fairtrade Bananas case study*. University of Greenwich: Natural Resources Insitute.

Bonnan-White, J., Hightower, A., and Issa, A. (2013). Of Couscous and Occupation: A Case Study of Women's Motivations to Join and Participate in Palestinian Fair Trade Cooperatives. *Agriculture and Human Values*, 30(3), pp.337-350.

Borot de Battisti, A., MacGregor, J., and Graffham, A. (2009). *Standard bearers: Horticultural exports and private standards in Africa*. London: IIED.

Bouroullec, M.D.M., and Paulillo, L.M. (2010). Governanças Híbridas no Comércio Justo Citrícola entre o Brasil e a Europa. In: APRESENTACAO ORAL-Estrutura, Evolução e Dinâmica dos Sistemas Agroalimentares e Cadeias Agroindustriais, pp.761-773.

Brenes, L., Piedra, B., Blanco, R., Salazar, D., Linares, J., and Rodriguez, A. (2011). Desarrollando un sistema de exportacion de pina organica basado en pequenos productores. *Acta horticulturae*, (906), pp.99-108.

Bruhn, C. (2012). *Canaan Fair Trade Impact Study 2005-2011*. Palestine: Canaan Fair Trade.

BSD Consulting. (2014). *Assessing the Benefits of Fairtrade Orange Juice for Brazilian Small Farmers*. São Paulo, Brazil: BSD Consulting.

Butler, F. (2014). Valuing unpaid labour in community Fair Trade products: a Nicaraguan case study from The Body Shop International. *Gender and Development*, 22(3), pp.533-547.

Calo, M., and Wise, T. (2005). *Revaluing Peasant Coffee Production: Organic and Fair Trade Markets in Mexico*. Medford, MA: Global Development and Environment Institute.

Cavino, V.C.O. (2012). *A experiência de comércio justo de agricultores familiares da cooperativa COOPFAM de Poço Fundo (MG): identificando os obstáculos e as potencialidades da transição para a agricultura de base ecológica.* Masters. Universidade Federal de São Carlos.

CEVAL (2012) *Fairtrade impact study: assessing the impact of Fairtrade on poverty reduction through rural development*. Center for Evaluation, University of the Saarland.

Chemnitz, C. (2007). The Compliance Decision with Food Quality Standards on Primary Producer Level. A Case Study of the EUREPGAP Standard in the Moroccan Tomato Sector. In: *I Mediterranean Conference of Agro-Food Social Scientists. 103rd EAAE Seminar "Adding Value to the Agro-Food Supply Chain in the Future Euromediterranean Space",* Barcelona, Spain.

Chemnitz, C. (2012). *The impact of food safety and quality standards on developing countries agricultural producers and exports*. Landwirtschaftlich-Gärtnerische Fakultät.

Chengappa, P.G., Rich, K.M., Muniyappa, A, Yadava, C.G., Ganashruthi, M.K., Pradeepa Babu, B.N., Shubha, Y.C., and Rich, M. (2014). *Sustainability coffee certification in India: Perceptions and Practices*. Norwegian Institute of International Affairs.

Chohin-Kuper, A., and Kemmoun, H. (2010). From theory to practice: fair trade in olive oil in Morocco. *Cahiers Agricultures*, 19, pp.17-22.

Colen, L., and Maertens, M. (2011). Private standards and employment insecurity: GlobalGAP in the Senegalese horticulture export sector. In: *EAAE Congress, Change and Uncertainty, Challenges for Agriculture, Food and Natural Resources,* Zurich, Switzerland.

Coles, C. (2011). Kilimanjaro and Oromia Coffee Value Chain Case Studies: Producer Benefits from Fair Trade and Free Market Channels. *NCCR North-South Dialogue*, 24.

Collinson, C., and Leon, M. (2000). *Economic Viability of Ethical Cocoa Trading in Ecuador*. Report 2519. University of Greenwich: Natural Resources Institute.

Collinson, C., Burnett, D., and Agreda, V. (2000). *Economic Viability of Brazil Nut Trading in Peru*. Report 2520. University of Greenwich: Natural Resources Institute.

COSA. (2013). *The COSA Measuring Sustainability Report: Coffee and Cocoa in 12 Countries.* Philadelphia, PA: The Committee on Sustainability Assessment.

COSA. (2013). *Vietnam Coffee: A COSA survey of UTZ certified farmers*. Philadelphia, PA: The Committee on Sustainability Assessment.

Coscione, M. (2013). Comercio Justo y Relevo Generacional. *AGO.USB Medellín-Colombia*, 13(2), pp.279-539.

Crespo, R. (2015). *Stream Friendly Coffee: Collaborating With The Rainforest Alliance To Advance Stream Conservation In Tropical Agroforestry Systems*. PhD. The University of Georgia.

Curtis, M. (2009). *The global food crisis and Fairtrade: small farmers, big solutions?* UK: The Fairtrade Foundation.

Dannenberg, P. (2011). Overcoming exclusion – informal dealing with the standard globalGAP In the Kenyan horticulture industry. *Geographische Zeitschrift*, 99(4), pp.237-255.

Dannenberg, P. (2011). Overcoming Exclusion-Informal Dealing with the Standard GlobalGAP in the Kenyan Horticulture Industry. *Geographische Zeitschrift*, 99(4), pp.237-255.

de Janvry, A., McIntosh, C., and Sadoulet, E. (2014). *Fair Trade and Free Entry: Can a Disequilibrium Market Serve as a Development Tool?* University of California at Berkeley.

De Luca, L.,Sahy, H., Joshi, S., Cortés, M. (2014). *Chapter 6: The Palestine Fair Trade Association (PFTA),occupied Palestinian territory.* In: Learning from Catalysts of Rural Transformation .Geneva, Switzerland: International Labour Organization.

de Mendonça, T.G. (2008). *Análise comparativa da viabilidade econômica da produção de mamão nos sistemas tradicional e integrada (PI) Comparative analysis of the economic feasibility of papaya production under the traditional and Integrated (IP) systems*. Masters. Universidade Federal de Viçosa.

de Oliveira, J.A.P., and Fortes, P.J.O.C. (2014). Global Value Chains and Social Upgrading of Clusters: Lessons from Two Cases of Fair Trade in the Brazilian Northeast. *Competition and Change*, 18(4), pp.365-381.

de Oliveira, R.F., Araújo, U.P., and dos Santos, A.C. (2008). Efeito do Fair Trade na Cooperativa de Agricultores Familiares de Café de Poço Fundo, MG *Organizaçōes Rurais & Agroindustriais, Lavras*, 10(2), pp.41-50.

Decisão Consultores. (2010). *Coffee Quality Project - Mogiana Region Brazil: Brazil Final Project Report*. São Paulo: Decisão Consultores.

Dengerink, J. (2013). *Improving livelihoods with private sustainability standards: measuring the development impact of the UTZ Certified*. Masters. Utrecht University.

Doherty, B., and Tranchell, S. (2005). New thinking in international trade? A case study of The Day Chocolate Company. *Sustainable Development*, 13(3), pp.166-176.

Dolan, C.S. (2008). In the Mists of Development: Fairtrade in Kenyan Tea Fields. *Globalizations*, 5(2), pp.305-318.

Dorr, A.C., Guse, J.C., Strassburger, R., Zulian, A., and Rossato, M.V. (2012). Economic analysis of the marketing channels in citrus industry in Brazil. *Annals of the University of Petrosani, and Economics*, 12(1), pp.93-108.

Duchelle, A.E. (2009). *Conservation and livelihood development in Brazil nut-producing communities in a tri-national Amazonian frontier*. PhD. University of Florida.

El Ouaamari, S. and Cochet, H. (2014). The Role of Coffee in the Development of Southwest Ethiopia's Forests: Farmers' Strategies, Investor Speculation, and Certification Projects. *Society & Natural Resources*, 27(2), pp.200-214.

Elder, S.D., Zerriffi, H., and Le Billon, P. (2012). Effects of fair trade certification on social capital: the case of Rwandan coffee producers. *World Development*, 40(11), pp.2355-2367.

Elder, S.D., Zerriffi, H., Le Billon, P. (2013). Is Fairtrade certification greening agricultural practices? An analysis of Fairtrade environmental standards in Rwanda. *Journal of Rural Studies*, (32), pp.264-274.

Elder, S.D.W. (2010). Fair Trade certification and social determinants of health: the case of coffee producers in Rwanda. Masters. The University of British Columbia.

Elfkih, S., Wannessi, O., and Mtimet, N. (2013). Le commerce equitable entre principes et réalisations: le cas du secteur oléicole Tunisien. *New Medit*, 12(1), pp.13-21.

Ellis, C.J. (2013). *"Como es duro es bonito" Labor Conditions and Gendered Complexities for Women Working on a Fairtrade Rose Farm in Ecuador*. Masters. University of California, Santa Barbara.

Enelow, N.H. (2012). *Fair Trade Coffee, Agrarian Cooperatives, and Rural Livelihoods in Northern Peru*. PhD. University of Massachusetts Amherst.

Ewert, J., Martin, A., and Nelson, V. (2005). *Assessment of the Social Impact of Corporate Codes of Practice In the South African Wine Industry*. University of Greenwich: Natural Resources Institute.

Fair Trade USA. (2011). *Cocoa Review 2010-11*. Oakland, California: Fair Trade USA.

Fair Trade USA. (2011). *Coffee Review 2010-11*. Oakland, California: Fair Trade USA.

Fair Trade USA. (2011). *Sugar Review 2010-11*. Oakland, California: Fair Trade USA.

Fairtrade Foundation. (2010). *Fairtrade Tea: Early Impacts in Malawi*. London: Fairtrade Foundation.

Fairtrade Foundation. (2010). *Impacts of Fairtrade in South Africa*. London: Fairtrade Foundation.

Fairtrade International. (2007). *The Benefits of Fairtrade - 2nd Edition: A Monitoring and Evaluation RPRT of Fairtrade Certified Producer Organizations for 2007*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2008). *The Benefits of Fairtrade - 2nd Edition: A Monitoring and Evaluation RPRT of Fairtrade Certified Producer Organizations for 2008*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2011). *Monitoring the Scope and Benefits of Fairtrade: Third Edition 2011*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2012). *Monitoring the Scope and Benefits of Fairtrade: Fourth Edition 2012*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2013). *Monitoring the Scope and Benefits of Fairtrade: Fifth Edition 2013*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2014). *Fairtrade Cocoa In West Africa*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2014). *Monitoring the Scope and Benefits of Fairtrade: Sixth Edition 2014*. Bonn, Germany: Fairtrade International.

Fairtrade International. (2014). *Unlocking the power together Annual Report 2013-2014*. Bonn, Germany: Fairtrade International.

Ferro-Soto, C., and Mili, S. (2013). Desarrollo rural e internacionalización. *Cuadernos de Desarrollo Rural*, 10(72), pp.267-289.

Frank, J., and Buckley, C.P. (2012). *Small-scale farmers and climate change: How can farmer organisations and Fairtrade build the adaptive capacity of smallholders?* London: IIED.

Fridell, G. (2005). *Fair trade in an unfair world? The prospects and limitations of social justice coffee, south and north*. PhD. York University.

Fridell, G. (2007). *Fair trade coffee: the prospects and pitfalls of market-driven social justice*. Toronto: University of Toronto Press.

Galizzi, P., and Herklotz, A. (2008). *The Role of the Environment in Poverty Alleviation*. New York: Fordham University Press.

Garcia, C., Ochoa, G., Garcia, J., Mora, J., and Castellanos, J. (2013). *Use of Polychoric Indexes to Measure the Impact of Seven Sustainability Programs on Coffee Growers' Livelihood In Colombia*. Manizales, Colombia: Centre for Regional Coffee and Entrepreneurial Research.

Garza, V., and Trejo, C. (2002). *Poverty Alleviation through participation in fair trade coffee networks: the case of Unión Majomut, Chiapas, Mexico.* San Cristóbal de Las Casas, Chiapas, México: Consultants for the Unión Majomut.

Geiger-Oneto, S., and Arnould, E.J. (2011). Alternative Trade Organization and Subjective Quality of Life: The Case of Latin American Coffee Producers. *Journal Of Macromarketing*, 31(3), pp.276-290.

Giovannucci, D., and Potts, J., with Killian, B., Wunderlich, C., Soto, G., Schuller, S., Pinard, F., Schroeder, K., and Vangeron, I. (2008). *Seeking Sustainability: COSA Preliminary Analysis of Sustainability Initiatives in the Coffee Sector*. Winnipeg, Canada: Committee on Sustainability Assessment.

Giovannucci, D., Brandriss, P., Brenes, E., Ruthenberg, I.M., and Agostini, P. (2000). *Engaging Civil Society to Create Sustainable Agricultural Systems: Environmentally-Friendly Coffee in El Salvador and Mexico.*

Gitter, S.R., Weber, J.G., Barham, B.L., Callenes, M., and Valentine, J.L. (2012). Fair Trade-Organic Coffee Cooperatives, Migration, and Secondary Schooling in Southern Mexico. *Journal of Development Studies*, 48(3), pp.445-463.

Gomes, C.C.M.P. (2010). *Alternativas encontradas para superação das principais dificuldades no processo de certificação Fair trade: um estudo multicasos de organizações de pequenos produtores no Brasil*. Masters. Universidade de São Paulo.

Gorritti, A.E.N. (2007). *Reflexions sur le commerce equitable au perou*. Masters. Université du Quebec a Rimouski.

Grieg-Gran, M. (2005). *From bean to cup: how consumer choice impacts upon coffee producers and the environment*. London: Consumers International.

Guedes, P.L. (2014). Group certification supports an increase in the diversity of SAN RA coffee producers in Brazil. *Ecological Economics*, 107, pp.59-64.

Guhl, A. (2009). Coffee, Forests, and Agricultural Certification in Aratoca, Santander. *Revista De Estudios Sociales*, (32), pp.114-125.

Guilherme, J.F. de A. (2011). *Análise das certificaçõ̃es agrícolas no pólo frutícola Petrolina (PE)/Juazeiro (BA*). Masters. Universidade Federal De Pernambuco.

Hafid, H., Neilson, J., Mount, T., and McKenzie, F. (2013). *Sustainability Impact Assessment of a certification scheme in the Indonesian cocoa industry: 2012 pilot survey results*. Sydney, Australia: University of Sydney.

Haynes, J., Cubbage, F., Mercer, E., and Sills, E. (2012). The Search for Value and Meaning in the Cocoa Supply Chain in Costa Rica. *Sustainability*, 4(7), pp.1466-1487.

Hellin, J., and Higman, S. (2002). Smallholders and niche markets: lessons from the Andes. *Network Paper - Agricultural Research and Extension Network*, 118.

Hidayat, N.K., Glasbergen, P., and Offermans, A. (2015). Sustainability Certification and Palm Oil Smallholders' Livelihood: A Comparison between Scheme Smallholders and Independent Smallholders in Indonesia. *International Food and Agribusiness Management Review*, 18(3), pp. 25-48.

Holzapfel, S., and Wollni, M. (2014). Is GlobalGAP Certification of Small-Scale Farmers Sustainable? Evidence from Thailand. *The Journal of Development Studies*, 50(5), pp. 731-747.

Hughell, D., and Newsom, D. (2013). *Impacts of Rainforest Alliance Certification on Coffee Farms in Colombia*. New York: Rainforest Alliance.

IIED and NRI. (2008). *Costs and benefits of EurepGAP compliance for African smallholders: A synthesis of surveys in three countries*. Fresh Insights Number 13. London: UK Department for International Development.

Imhof, S., and Lee, A. (2007). *Assessing the Potential of Fair Trade for Poverty Reduction and Conflict Prevention: A Case Study of Bolivian Coffee Producers*. Cape Town, South Africa: Fairtrade South Africa.

Imke, H., Girardo, C., and Converso, D. (2011). La traducción de los valores democráticos en la participación organizativa de la sociedad civil: un estudios de caso entre México e Italia. *Economía, Sociedad y Territorio,* (11)37, pp.667-706.

Ingram, V., Waarts, Y., Ge, L., van Vugt, S., Wegner, L., Puister-Jansen, L., Ruf, F., and Tanoh, R. (2014). *Impact of UTZ certification of cocoa in Ivory Coast: Assessment framework and baseline*. LEI Wageningen UR.

Ingram, V., Waarts, Y., van Vugt, S.M., Ge, L., Wegner, L., and Puister-Jansen L. (2014). *The impact of UTZ certification of cocoa in the Ivory Coast 2008 to 2013*. LEI Wageningen UR.

Jaffee, D. (2008). 'Better, but not great': the social and environmental benefits and limitations of Fair Trade for indigenous coffee producers in Oaxaca, Mexico. In: Ruben, R., *The Impact of Fair Trade.* Wageningen: Wageningen Academic Publishers, pp.195-220.

Johannesen, S. (2008). *Producing and Consuming Narratives: The Value of Fairtrade Coffee*. Master. Norwegian University of Life Sciences.

Johannessen, S., and Wilhite, H. (2010). Who really benefits from Fairtrade? An analysis of value distribution in Fairtrade coffee. *Globalizations*, 7(4), pp.525-544.

Jones, E., Smith, S., and Wills, C. (2012). Women producers and the benefits of collective forms of enterprise. *Gender and Development*, 20(1), pp.13-32.

Jordaan, H., and Grove, B. (2013). Transaction cost analysis of raisins marketing by emerging farmers from Eksteenskuil, South Africa. *Agrekon*, 52(4), pp.21-42.

Karira, W.J., Mburu, J., and Guthiga, P.M. (2013). Empirical Analysis of the Environmental Benefits of Compliance with GLOBALGAP Standards Among Smallholder Farmers in Eastern and Central Kenya: An Environmental Valuation Approach. In: *4th International Conference of the African Association of Agricultural Economists,* Hammamet, Tunisia.

Kariuki, I.M. (2014). Transition to certification schemes and implications for market access: GlobalGAP perspectives in Kenya. *Agricultural Sciences*, 5(12), pp.1100-1111.

Kariuki, I.M., Loy, J.P., and Herzfeld, T. (2012). Farmgate Private Standards and Price Premium: Evidence from the GlobalGAP Scheme in Kenya's French Beans Marketing. *Agribusiness*, 28(1), pp.42-53.

Kiemen, A., and Beuchelt, T. (2010). Certification as an upgrading strategy for small-scale farmers and their cooperatives : a value chain analysis for Nicaraguan coffee. *Research in Development Economics and Policy (Discussion Paper Series)*, (2/2010), pp.570-582.

Kilian, B., Pratt, L., Jones, C., and Villalobos, A. (2004). Can the Private Sector be Competitive and Contribute to Development through Sustainable Agricultural Business? A Case Study of Coffee in Latin America. *International Food and Agribusiness Management Review*, 7(3), pp.375-386.

Kirumba, E.G., and Pinard, F. (2010). Determinants of farmers' compliance with coffee eco-certification standards in Mt. Kenya region. In: *Joint 3rd African Association of Agricultural Economists (AAAE) and 48th Agricultural Economists Association of South Africa (AEASA) Conference,* Cape Town, South Africa.

Kleine, D. (2008). Negotiating partnerships, understanding power: doing action research on Chilean Fairtrade wine value chains. *Geographical Journal*, 174(2), pp.109-123.

Kodama, Y. (2009). The Effects of Fair Trade on Coffee Producers: A Case Study of Ethiopian Coffee Cooperatives. In: *Proceedings of the 16th International Conference of Ethiopian Studies*, Trondheim, Norway, pp. 779-792.

Krishna, V.V., Pascual, U., and Zilberman, D. (2010). Assessing the Potential of Labelling Schemes for In Situ Landrace Conservation: An Example from India. *Environment and Development Economics*, 15(2), pp.127-151.

Kruger, S. (2008). *A case study of Fairtrade labelling and worker empowerment on two wine and fruit farms in the Western Cape.* Masters. University of Western Cape.

Krupka, J. (2012). *The Fair Trade Coffee Business Model's Affect on the Small Scale Producers through the Lens of the Triple Bottom Line*. PhD. Georgia State University.

Kuwornu, J.K.M., and Mustapha, S. (2013). Global GAP standard compliance and smallholder pineapple farmers' access to export markets: implications for incomes. *Journal of Economics and Behavioral Studies*, 5(2), pp.69-81.

Lalitha, N., Nelson, V., Martin, A., and Posthumus, H. (2013). *Final Technical Report: Assessing the poverty impact of sustainability standards: Indian tea*. University of Greenwich: Natural Resources Unit.

Laroche, K., Jiménez, R., and Nelson, V. (2012). *Assessing the Impact of Fairtrade for Peruvian Cocoa Farmers*. University of Greenwich: Natural Resources Unit.

Lazaro, E., Makindara, J., and Kilima, F. (2008). *Sustainability Standards and Coffee Exports From Tanzania*. Copenhagen, Denmark: Danish Institute for International Studies.

Lemeilleur, S. (2012). *International standards and small-scale farmer behaviors: evidence from Peru*. In: *VIII International Agribusiness PAA-PENSA Conference "The Multiple Agro Profiles: How to Balance Economy, Environment and Society",* Buenos Aires, Argentina.

Lewis, J., and Runsten, D. (2008). Is Fair Trade-Organic Coffee Sustainable in the Face of Migration? Evidence from a Oaxacan Community. *Globalizations*, 5(2), pp.275-290.

Lo Turco Martinez, J.R., Jesus, J.C.D.S., and Cócaro, H. (2008). "Casos Sobre A Certificação Utz Kapeh Em Empresas Cafeeiras Informatizadas: Impactos Nas Pessoas, Gestão E Competitividade. In: *Sociedade Brasileira de Economia, Administracao e Sociologia Rural (SOBER), 46th Congress*, Rio Branco, Acre, Brasil.

Loconto, A. (2010). Value Chains and Chains of Values: Tracing Tanzanian Tea. In: *European Association of Agricultural Economists>116th Seminar*, Parma, Italy, pp.134-145.

Loconto, A.M. (2010). *Sustainabilitea: Shaping sustainability in Tanzanian tea production*. PhD. Michigan State University.

Luetchford, P. (2008). The hands that pick fair trade coffee: beyond the charms of the family farm. In: De Neve, G., Luetchford, P., Pratt, J., and Wood, D.C., *Hidden Hands in the Market: Ethnographies of Fair Trade, Ethical Consumption, and*

Corporate Social Responsibility, Bingley, UK: Emerald Group Publishing, pp.143-170.

Luna, F., and Wilson, P.N. (2015). An Economic Exploration of Smallholder Value Chains: Coffee Transactions in Chiapas, Mexico. *International Food and Agribusiness Management Review*, (18)3, pp. 85-106.

Luvai, L. (2008). *Private standard impacts on developing country producers: a personal experience of GlobalGap certification in Kenya*. Fresh Perspectives, Agrifood Standards.

Lyon, S. (2002). *Evaluation of the Actual and Potential Benefits for the Alleviation of Poverty Through the Participation in Fair Trade Coffee Networks: Guatemalan Case Study*.

Lyon, S. (2007). Maya coffee farmers and fair trade: assessing the benefits and limitations of alternative markets. *Culture & Agriculture*, 29(2), pp.100-112.

Macawaris, N., Taylor, C., and Zamora-Galinato, C. (2011). *Sowing the Seeds of Sustainability: A Case Project with Unifrutti, IFC, and Smallholder Banana Farmers in the Philippines.* Washington, DC: International Finance Corporation.

Madjidi, O. (2011). *Sustainable coffee certification programs and coffee cooperatives in Guatemala: A small-scale producer perspective*. Masters. Royal Roads University.

Malini, H., and Aryani, D. (2012). Efficiency analysis of palm plantation plasma farmers certified RSPO and without RSPO on sustainable palm plantation management in Musi Banyuasin Regency South Sumatera Indonesia. In: *International Proceedings of Chemical, and Biological and Environmental Engineering (IPCBEE)*, pp.194-199.

Malins, A., and Blowfield, M. (2000). *Fruits of the Nile, Fairtrade Processing case study.* University of Greenwich: Natural Resources Institute.

Malins, A., and Nelson, V. (1998). *Ethical trade and sustainable rural livelihoods - case studies: Farmers Fair Trade (Uganda) Ltd, Organic cotton trade case study*. University of Greenwich: Natural Resources Institute.

Martinez, M. (2002). *Poverty Alleviation Through Participation in Fair Trade Coffee Networks: The Case of the Tzotzilotic Coffee Coop Chiapas, Mexico.* Mexico: Desarrollo Alternativo, A.C.

McEwan, C., and Bek, D. (2009). Placing ethical trade in context: wieta and the South African wine industry. *Third World Quarterly*, 30(4), pp.723-742.

McEwan, C., and Bek, D. (2009). The political economy of alternative trade: Social and environmental certification in the South African wine industry. *Journal of Rural Studies*, 25, pp.255-266.

McEwan, C., Hughes, A., Bek, D., and Rosenberg, Z. (2014). Why 'place' matters in the development and impacts of Fairtrade production. *Food Chain*, 4(1), pp.78-82.

Mendez, E., Bacon, C., Olson, M., Petchers, S., Herrador, D., Carranza, C., Trujillo, L., Guadarrama-Zugasti, C., Cordo, A., and Mendoza, A. (2010). Effects of Fair Trade and organic certifications on small-scale coffee farmer households in Central America and Mexico. *Renewable Agriculture and Food Systems,* 25(3), pp.236-251.

Meneley, A. (2011). Blood, sweat and tears in a bottle of Palestinian extra-virgin olive oil. *Food, Culture, and Society*, 14(2), pp.275-292.

Meneley, A. (2014). Discourses of distinction in contemporary Palestinian extra-virgin olive oil production. *Food and Foodways*, 22(1-2), pp.48-64.

Meskela, T., and Teshome, Y. (2014). From economic vulnerability to sustainable livelihoods: the case of the Oromia Coffee Farmers Cooperatives Union (OCFCU) in Ethiopia. *International Food and Agribusiness Management Review*, 17B(Special Issue), pp.103-108.

Milford, A.B. (2012). The Pro-Competitive Effect of Coffee Cooperatives in Chiapas, Mexico. *Journal of Agricultural & Food Industrial Organization*, 10(1).

Miller, K. (2009). *Fair Trade Fund Support of Nicaraguan Coffee Farmers: A Case Study*. Catholic Relief Services.

Monsarrat, M.T. (2002). *Fair trade and awareness on the ground: The case of CECOCAFEN in Nicaragua*. Masters. Simon Fraser University.

Mori, T., and Malo, M. (2003). Impactos del comercio justo del vino. Tres casos de empresas colectivas de productores en Chile. *CIRIEC-Espana, Revista de Economia Publica, and Social y Cooperativa*, (46), pp.265-289.

Morse, P. (2004). *The fair trade movement and globalization: No one need die marching from Matagalpa to Managua*. Masters. University of Massachusetts Lowell.

Moseley, W.G. (2008). Fair Trade Wine: South Africa's Post-Apartheid Vineyards and the Global Economy. *Globalizations*, 5(2), pp.291-304.

Mugabe, T. (2011). *Fair Trade in the Eastern Cape: an examination of its socio-economic impact and challenges among emerging black farmers*. Masters. University of Fort Hare, Faculty of Social Sciences and Humanities.

Muller, C., Vermeulen, W.J.V., and Glasbergen, P. (2009). Perceptions on the demand side and realities on the supply side: a study of the South African table grape export industry. *Sustainable Development*, 17(5), pp.295-310.

Muriithi, B. (2008). *Compliance With Eurepgap Standards: Determinants, Costs and Implications on Profitability Among Small Holder French Beans Exporters in Kirinyaga District, Kenya*. Masters. Egerton University.

Murray, D., Raynolds, L., and Taylor, P. (2006). The Future of Fair Trade Coffee: Dilemmas Facing Latin America's Small-Scale Producers. *Development in Practice*, 16(2), pp.179-192.

Murray, D., Raynolds, L.T., and Taylor, P.L. (2003). *One cup at a time - poverty alleviation and Fair Trade in Latin America*. Degree. Colorado State University.

Napitupulu, T.A., Natawidjaja, R.S., and Lantican, J.M. (2009). Adding Value to Fresh and Processed Produce through Product Certification. *CAPSA Working Paper No. 104*.

Narvaez, G.E. (2010). *Solidarity, sustainability and standards: US-Nicaragua specialty coffee networks*. PhD. University of California, Irvine.

Nel, E., Binns, T., and Bek, D. (2007). 'Alternative foods' and community-based development: Rooibos tea production in South Africa's West Coast Mountains. *Applied Geography*, 27(2), pp.112-129.

Nelson, V., and Galvez, M. (2000). *Social Impact of Ethical and Conventional Cocoa Trading on Forest-Dependent People in Ecuador.* University of Greenwich: Natural Resources Institute.

Nelson, V., and Malins, A. (1998). *Impact assessment of a fair trade, organic cotton scheme, Uganda*. University of Greenwich: Natural Resources Institute.

Nelson, V., and Smith, S. (2011). *Fairtrade Cotton: Assessing Impact in Mali, Senegal, Cameroon and India.* University of Greenwich: Natural Resources Institute.

Nicetic, O., van de Fliert, E., van Chien, H., Mai, V., and Cuong, L. (2010). Good Agricultural Practice (GAP) as a vehicle for transformation to sustainable citrus production in the Mekong Delta of Vietnam. In: *Building sustainable rural futures: the added value of systems approaches in times of change and uncertainty. 9th European IFSA Symposium,* Vienna, Austria, pp.1893-1901.

Ochieng, B.O., Hughey, K.F.D., and Bigsby, H. (2013). Rainforest Alliance Certification of Kenyan tea farms: a contribution to sustainability or tokenism? *Journal of Cleaner Production*, 39, Pp.285-293.

Odegard, R.M. (2014). *Fair Trade in Transition: Evolution, Popular Discourse, and the Case of the CADO Cooperative in Cotopaxi, Ecuador*. Masters. University of South Florida.

Okello, J.J., Narrod, C.A., and Roy, D. (2007). *Institutional Innovations for Smallholder Compliance with International Food Safety Standards: Experiences from Kenya, Ethiopian and Zambian Green Bean Growers.* In: African Association of Agricultural Economists (AAAE)> Second International Conference, Accra, Ghana, pp.559-571.

Omosa, M., Kimani, M., and Njiru, R. (2006). *Assessment of the Social Impact of Corporate Codes of Practice in the Kenyan Cut Flower Industry*. University of Greenwich: Natural Resources Institute.

Oseni, J.O., and Adams, A.Q. (2013). Cost Benefit Analysis of Certified Cocoa Production in Ondo State, Nigeria. In: African Association of Agricultural Economists (AAAE) Fourth International Conference, Hammamet, Tunisia. pp.584-599.

Osertag, C., Sandoval, O., Barona, J, and Mancilla, C. (2014). *An Evaluation of Fairtrade Impact on Smallholders and Workers in the Banana Sector in Northern Colombia - Final Report*. The Netherlands: The Corporation for Rural Business Development.

Otieno, G., and Knorringa, P. (2012). Localizing Global Standards: Illustrative Examples from Kenya's Horticulture Sector. In: van Dijk, M.P., and Trienekens, J., *Global Value Chains: Linking Local Producers from Developing Countries to International Markets,* Amsterdam: Amsterdam University Press, pp.119-135.

Owusu-Amankwah, R., Ruivenkamp, G., Essegbey, G., and Frempong, G. (2014). Implications of third party voluntary cocoa certification on labour and livelihood systems in Ghana. *International Journal of Agriculture Innovations and Research*, 2(6), pp.1047-1059.

Palmieri, R.H. (2008). *Impactos socioambientais da certificação Rainforest Alliance em fazendas produtoras de café no Brasil Socio-environmental impacts of Rainforest Alliance Certification on Brazilian coffee plantations*. Masters. Universidade de São Paulo.

Parrish, B.D., Luzadis, V.A., and Bentley, W.R. (2005). What Tanzania's Coffee Farmers Can Teach the World: A Performance-Based Look at the Fair Trade-Free Trade Debate. *Sustainable Development*, 13, pp.177-189.

Partner Africa. (2013). *Waitrose Supply Chain Impact Assessment: Tekangu Cooperative, Kenya*. London: Carnstone Partners LLP.

Paschall, M. (2013). *The Role of Third Party Certification in Improving Small Farmer Livelihoods.* PhD. University of St. Gallen.

Pemberton, M. (2010). *Alternative, unconventional or business-as-usual? Implications and experiences of Fairtrade producers in the free market*. Masters. University of Guelph.

Perez-Aleman, P. (2013). Regulation in the Process of Building Capabilities: Strengthening Competitiveness While Improving Food Safety and Environmental Sustainability in Nicaragua. *Politics & Society*, 41(4), pp.589-620.

Philpott, S.M., Bichier, P., Rice, R., and Greenberg, R. (2007). Field-testing ecological and economic benefits of coffee certification programs. *Conservation Biology*, 21(4), pp.975-985.

Pinto, L.F.G., Gardner T., McDermott, C.L., and Ayub, K.O.L. (2014). Group Certification Supports an Increase in the Diversity of Sustainable Agriculture Network-Rainforest Alliance Certified Coffee Producers in Brazil. *Ecological Economics*, 107, pp.59-64.

Poncelet, M., Defourny, J., and De Pelsmaker, P. (2005). *Fair and Sustainable Trade, between Markt and Solidarity: Diagnosis and Prospects*. Brussels, Belgium: Belgian Science Policy.

Pongratz-Chander, K. (2007). *Fair trade coffee in Costa Rica: A new model for sustainable development?* PhD. Northern Arizona University.

Potts, J. (2007). *Alternative trade initiatives and income predictability: theory and evidence from the coffee sector.* Winnipeg, Manitoba: International Institute for Sustainable Development.

Pound, B., and Phiri, A. (2011). *Taking Root: Fairtrade in Malawi*. London: Fairtrade Foundation.

Powae, W.I. (2009). *Fair trade coffee supply chains in the highlands of Papua New Guinea: do they give higher returns to smallholders?* Masters. Lincoln University.

Rainforest Alliance. (2012). *Redesigning Land-Use and Business Practices: 25 Years of Impacts*. New York: Rainforest Alliance.

Rainforest Alliance. (2013). *Farmer Bankability and Sustainable Finance Farm-Level Metrics that Matter*. New York: Rainforest Alliance.

Rainforest Alliance. (2013). *Investing in Sustainability and Productivity Improvements to Transform Cocoa Production and Livelihoods in Côte d'Ivoire*. New York: Rainforest Alliance.

Ramm, G., Künkel, P., and Fricke, V. (2008). *Introduction of Voluntary Social and Ecological Standards in Developing Countries: Synthesis Report*. Bonn: Federal Ministry for Economic Cooperation and Development.

Rebelo, P. (2013). *How can a certification process focused on GAP provide better access to international markets for farmers?* Wageningen UR: Centre for Development Innovation.

Reck, S. (2010). *Investigating the Interplay of Producer Organisations and Fairtrade Certification and their Impact on Farmers' Livelihood Capitals: A Case Study of the Fairtrade Certified Producer Organisation ASOPROAGROIN in Northern Costa Rica*. MSc. Universiteit Van Amsterdam.

Renner, M., and Adamowicz, W. (1998). *The Effects of Alternative Agriculture and Fair Trade on the Development of Producer Groups and Their Members: Case Studies from Northern Thailand*. Staff Paper 98-06. University of Alberta: Department of Rural Economy.

Renner, M.E. (1998). *Measuring the effects of alternative agricultural methods and fair trade practices on the development of producer groups and their members in northern Thailand*. Masters. University of Alberta.

Robins, N., Roberts, S., and Abbot, J. (2000). *Who Benefits? A social assessment of environmentally-driven trade*. London: IIED.

Robinson, P. (2010). *Do voluntary labour initiatives make a difference for the conditions of workers in global supply chains? A case study of the Costa-Rican United Kingdom banana chain*. Cardiff University: The ESRC Centre for Business Relationships, Accountability, Sustainability and Society.

Romanoff, S. (2010). Shade coffee in biological corridors: potential results at the landscape level in El Salvador. *Culture and Agriculture*, 32(1), pp.27-41.

Ruben, R., and van Schendel, L. (2008). The impact of Fair Trade on banana plantations in Ghana: income, ownership and livelihoods of banana workers. In: Ruben, R., *The Impact of Fair Trade.* Wageningen: Wageningen Academic Publishers, pp.137-153.

Ruben, R., Clercx, L., Cepeda, D., and de Hoop, T. (2008). Fair Trade impact of banana production in El Guabo Association, Ecuador: a production function analysis. In: Ruben, R. (Ed.) *The Impact of Fair Trade.* Wageningen: Wageningen Academic Publisher, pp.155-166.

Ruben, R., Fort, R., and Zúñiga-Arias, G. (2009). Measuring the impact of fair trade on development. *Development in Practice*, 19(6), pp.777-788.

Rueda, X., and Lambin, E.F. (2013). Linking Globalization to Local Land Uses: How Eco-Consumers and Gourmands are Changing the Colombian Coffee Landscapes. *World Development*, 41, pp.286–301.

Samnegård, E. (2007). *How Does Fairtrade Affect the Market? A Case Study in South Africa.* MSc. Lund University.

Sánchez, A.M., Águila González, J.M., Pérez Poumián, M.L., Cruz Castillo, R.deJ. (2013). El Control Interno de una Organización productora de Café Certificado, en Chiapas, Mexico. *Revista Mexicana de Agronegocios*, 17(33), pp.460-470.

Saysana, V. (2011). *Promoting organic and fair trade certification in the Lao PDR coffee sector: Benefits and challenges for farmers and local economies*. Masters. Oklahoma State University.

Schreinemachers, P., Schad, I., Tipraqsa, P., Williams, P.M., Neef, A., Riwthong, S., Sangchan, W., and Grovermann, C. (2012). Can public GAP standards reduce agricultural pesticide use? The case of fruit and vegetable farming in northern Thailand. *Agriculture And Human Values*, 29(4), pp.519-529.

Schut, M., and Florin, M.J. (2015). The policy and practice of sustainable biofuels: Between global frameworks and local heterogeneity. The case of food security in Mozambique. *Biomass & Bioenergy*, 72, pp.123-135.

Schweisguth, M.A. (2015). *Evaluating the Effects of Certification on Smallholders? Net Incomes, with a Focus on Cacao Farmers in Cooperatives in Cote d'Ivoire*. Masters. University of California, Davis.

Selfa, T., Bain, C., and Moreno, R. (2014). Depoliticizing land and water “grabs” in Colombia: the limits of Bonsucro certification for enhancing sustainable biofuel practices. *Agriculture and Human Values*, 31(3), pp.455-468.

Sen, D. (2014). Fair trade vs. Swaccha Vyapar: women's activism and transnational justice regimes in Darjeeling, India. *Feminist Studies*, 40(2), pp.444-472.

Shankar, F.E. (1999). *Organic beans and ethical aromas: Small-scale coffee production in Guatemala*. Masters. Queen's University.

Shanley, P., Pierce, A., Laird, S., and Robinson, D. (2008). *Beyond Timber: Certification and management of non-timber forest products*. Bogor Barat, Indonesia: Center for International Forestry Research.

Shreck, A. (2002). Just bananas? Fair Trade banana production in the Dominican Republic. *International Journal of Sociology of Agriculture and Food*, 10(2), pp.13-23.

Sick, D. (2008). Coffee, farming families, and fair trade in Costa Rica: new markets, same old problems? *Latin American Research Review*, 43(3), pp.193-208.

Silva, E. (2012). *Fatores determinantes da adoção de certificações*. Masters. Universidade Federal de Lavras.

Silva, E.C., Castro Junior, L.G., Costa, C.H.G., and Andrade, F.T. (2014). Determinant factors in adopting socio-environmental certifications in coffee farms. *Revista de Economia e Sociologia Rural*, 52(3), pp.437-448.

Smith, A.M. (2013). Fair trade governance and diversification: the experience of the National Smallholder Farmers' association of Malawi. *Geoforum*, 48, pp.114-125.

Smith, K. (2012). Gender and Food Security in a Fair, Green Economy? *Development*, 55(1), pp.81-89.

Stellmacher, T. (2008). *Prospects and challenges of forest coffee certification in Ethiopia: the need to effectively link economic benefits and biodiversity conservation*. Bonn, Germany: Center for Development Research.

Takahashi, R., and Todo, Y. (2014). The impact of a shade coffee certification program on forest conservation using remote sensing and household data. *Environmental Impact Assessment Review*, 44, pp.76-81.

Tallontire, A. (2000). Partnerships in fair trade: reflections from a case study of Cafédirect. *Development in Practice*, 10(2), pp.166-177.

Tallontire, A., Smith, S., and Njobvu, C. (2004). *Ethical trade in African horticulture: gender, rights and participation*. University of Greenwich: National Resources Institute.

Tiffen, P. (2002). A Chocolate-Coated Case for Alternative International Business Models. *Development in Practice*, 12(3-4), pp. 383-397.

Torgerson, A. (2007). *Fair Trade and Global Justice: The Radical Possibilities of Reform*. MA. Carleton University.

Torgerson, A.M. (2010). Fair trade banana production in the Windward Islands: local survival and global resistance. *Agriculture and Human Values*, 27(4), pp.475-487.

TRAIDCRAFT. (2014). *Development Report*. Tyne and Wear, UK: TRAIDCRAFT.

TRAIDCRAFT. (2014). *Impact and performance Report*. Tyne and Wear, UK: TRAIDCRAFT.

TWIN. (2013). *Empowering Women Farmers In Agricultural Value Chains*. London: TWIN.

TWIN. (2014). *Development Through Trade: Annual Report 2013-2014*. London: TWIN.

Udomkit, N. (2002). *Three dimensions of fair trade: The analysis of financial, social, and environmental objectives of fair trade in organic rice*. . University of Bath.

Utting-Chamorro, K. (2005). Does Fair Trade make a difference? The case of small coffee producers in Nicaragua. *Development in Practice*, 15(3/4), pp.584-599.

UTZ. (2016). *UTZ Impact Report 2016*. Amsterdam, The Netherlands: UTZ.

Vagneron, I., and Roquigny, S. (2011). Value distribution in conventional, organic and fair trade banana chains in the Dominican Republic. *Revue canadienne d'études du développement*, 32(3), pp.324-338.

Valentini, E. (2010). Characteristics and organization of the Fairtrade market in Brazil. In: 116th EAAE SEMINAR "Spatial Dynamics in Agri-food Systems: Implications for Sustainability and Consumer Welfare". Parma, Italy.

Valkila, J. (2009). Fair Trade Coffee Certification: A tool for rural development and environmental protection in Nicaragua?. *Department of Economics and Management Discussion Papers*, 38, pp..

Valkila, J. (2014). Do Fair Trade pricing policies reduce inequalities in coffee production and trade? *Development Policy Review*, 32(4), pp.475-493.

Valkila, J., and Nygren, A. (2008). Impacts of Fair Trade-certification on Coffee Farmers, Cooperatives and Laborers, in Nicaragua. *Agriculture and Human Values*, 27(3), pp.321-333.

Valkila, J., Haaparanta, P., and Niemi, N. (2010). Empowering Coffee Traders? The Coffee Value Chain from Nicaraguan Fair Trade Farmers to Finnish Consumers. *Journal of Business Ethics*, 97(2), pp.257-270.

van der Waal, J.W.H., and Zongo, A. (2011). Developing a fresh mango export value chain with West-African smallholder mango farmers. *Acta Horticulturae*, (895), pp. 283-291.

Vieira, L.M., and Maia, T. (2009). The governance of fair trade system: evidence from small honey producers in Rio Grande do Sul. *Brazilian Administrative Review*, (6)4, pp.367-379.

Villegas, M. (2004). *Financial viability of eco-labeled cacao in the Santander Mountains (Colombia)*. Masters. San Jose State University.

Waarts, Y., Lan, G., Giel, T., and van der Mheen, J. (2013). *A touch of cocoa: Baseline study of six UTZ- Solidaridad.* LEI Wageningen UR.

Waridel, L. (2002). *Sustainable trade: The case of coffee in North America*. Masters. University of Victoria.

WRENmedia. (2007). *Voices from the field: Published articles and case studies. Bridging the GAP, supporting smallholders in Kenyan export horticulture.* Eye, UK: WRENmedia.

Yamoah, F., O'Caoimh, C., Donnelly, C., and Sawaya, S.K. (2014). The Journey from Subsistence to Commercial Viability: The Case of Meru Herbs, Kenya. *International Food and Agribusiness Management Review*, 17, pp.139-144.

Additional References

Aasprong, H. (2013). Entangled standardizing networks: the case of GLOBALGAP and Fairtrade in St Vincent's banana industry. *International Journal of Sociology of Agriculture and Food*, 20 (1), 91-108.

Barratt Brown, M. (1993). *Fair Trade*. London: Zed Books.

Barrientos, S. (2000). Ethical trade and globalisation: assessing the implications for development. *Journal of International Development*, 12(4), pp.559-570.

Barrientos, S., and Smith, S. (2006). *The ETI Code of Labour Practice: Do Workers Really Benefit? Report on the ETI Impact Assessment 2006. Part 4: Research Methodology*. Brighton: Institute of Development Studies, University of Sussex.

Barrientos, S., Dolan, C., and Tallontire, A. (2003). A gendered value chain approach to codes of conduct in African horticulture. *World Development*, 31(9), pp.1511–1526.

Bennett, M., and Franzel, S. (2013). Can organic and resource-conserving agriculture improve livelihoods? A synthesis. *International Journal of Agricultural Sustainability*, 11(3), pp.193-215.

Bhowmik, S. K. (1980). The plantation as a social system. *Economic and Political Weekly*, 15(36), pp.1524-1527.

Blackman, A., and Rivera, J. (2010). *The evidence base for environmental and socioeconomic impacts of "sustainable" certification*. Washington D.C.: Resources for the Future.

Blowfield, M. (1999). Ethical trade: a review of developments and issues. *Third World Quarterly*, 20(4), pp.753- 770.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons.

CBI. (2005). *Fresh Fruit and Vegetables: EU Market Survey 2005*. The Hague: Centre for the Promotion of Imports from Developing Countries.

Chan, M.K. and Pound, B. (2009). *Final report: literature review of sustainability standards and their poverty impact.* NRI, London.

Chiputwa, B., Spielman, D. J., and Qaim, M. (2015). Food standards, certification, and poverty among coffee farmers in Uganda. *World Development*, 66, pp.400-412.

Critical Appraisal Skills Programme (CASP). (2013). *10 questions to help you make sense of qualitative research*. Public Health Resource Unit: England. Retrieved from: www.phru.nhs.uk/Doc_Links/Qualitative%20Appraisal%20Tool.pdf

Committee on Sustainability Assessment (COSA) and Institute of Statistical Social and Economic Research (ISSER). (2013). *Cocoa Farms in Ghana: An Evaluation of the impact of UTZ Certification on the Sustainability of Smallholders Supported by the Solidaridad Cocoa Programme (2010-2012).*

Daviron B. and Gibbon, P. (2002). 'Global commodity chains and African export agriculture'. *Journal of Agrarian Change,* 2(2), pp.137-161.

Daviron, B., and Ponte, S. (2005). *The coffee paradox: global markets, commodity trade and the elusive promise of development.* New York: Zed Books.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton, NJ: Princeton University Press.

Dickersin, K. (2006). Publication bias: recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein, H. R., Sutton, A. J., and

Borenstein, M. (Eds.). *Publication bias in meta-analysis: prevention, assessment and adjustments*. Chichester: John Wiley & Sons.

Doherty, B., and Tranchell, S. (2005). New thinking in international trade? A case study of The Day Chocolate Company. *Sustainable Development*, 13(3), 166-176.

Dolan, C., 2004. 'On Farm and Packhouse: Employment at the Bottom of a Global Commodity Chain'. *Rural Sociology* 69(1), pp.99–126.

Dragusanu, R., Giovannucci, D., and Nunn, N. (2014). The economics of fair trade. *Journal of Economic Perspectives*, *28*(3), pp.217-236.

Duvendack , M., Hombrados, J. G., Palmer-Jones, R. and Waddington, H. (2012). Assessing 'what works' in international development: meta-analysis for sophisticated dummies. *Journal of Development Effectiveness*, 4(3), pp.456-471.

Ellis, P. D. (2010) *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*, Cambridge: Cambridge University Press.

Fairtrade Foundation. (2012). *Fairtrade and coffee: commodity briefing*. London: Fairtrade Foundation.

Fairtrade Foundation. (2013). *Powering up smallholder farmers to make food fair: A five point agenda*. London: Fairtrade Foundation.

Fairtrade Foundation. (2014). *Annual impact report 2013-2014*. London: Fairtrade Foundation.

Fairtrade International (2016). *Monitoring the scope and benefits of Fairtrade Seventh Edition, 2015*. Bonn : Fairtrade International.

Fairtrade International, (2017). *Journeys to change: Fairtrade theory of change.* [online]. Accessed on January 10, 2017. Available at: http://www.fairtrade.net/fileadmin/user_upload/content/2009/resources/1612-Fairtrade_Theory_of_Change.pdf

Fairtrade, Employment and Poverty Reduction team (FTEPR), 2014. *Fairtrade, Employment and Poverty Reduction in Ethiopia and Uganda*. Final Report to DFID, April 2014. London: SOAS, University of London. Retrieved from http://ftepr.org/publications/.

Geertz, C. (1973) Thick description: toward an interpretative theory of culture. In: Geertz, C. *The interpretation of culture: selected essays*. New York: Basic Books.

Gereffi, G., and M. Korzeniewicz (1994). *Commodity chains and global capitalism*. ABC-CLIO.

Gereffi, G., Garcia-Johnson, R., and Sasser, E. (2001). The NGO-industrial complex. *Foreign Policy*, 56-65.

Ghosh, J. (2010). The unnatural coupling: Food and global finance. *Journal of Agrarian Change*, 10(1), pp.72-86.

Gibbon, P., and Ponte, S. (2005). *Trading down: Africa, value chains, and the global economy*. Philadelphia PA: Temple University Press.

GlobalGAP (2016). Who we are. What we do. http://www.globalgap.org/uk_en/what-we-do/

Greenhalgh, T., and Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *British Medical Journal*, 331(7524), pp.1064-1065.

Hallam, D., et al. (2004). *The market for non-traditional agricultural exports*. Rome: FAO.

Hammerstrøm, K., Wade, A., and Klint Jørgensen, A.M. (2010). *Searching for studies: a guide to information retrieval for Campbell Systematic Reviews*. Campbell Collaboration: Oslo.

Hansen, H., and Trifkovi□, N. (2014). Food Standards are good–for middle-class farmers. *World Development*, 56, pp.226-242.

Henson, S. and Humphrey, J. (2010). Understanding the complexities of private standards in global agri-food chains as they impact developing countries. *Journal of Development Studies,* 46, pp.1628–1646.

Henson, S., and Jaffee, S. (2008). Understanding developing country strategic responses to the enhancement of food safety standards. *The World Economy*, 31(4), pp.548-568.

Hickey, S., and Du Toit, A. (2013). 'Adverse incorporation, social exclusion, and chronic poverty'. In *Chronic Poverty*, Palgrave Macmillan UK, pp.134-159.

Hoebink, P., Ruben, R., Elbers, W., and van Rijsbergen, B., ed. 2014, *The impact of coffee certification on smallholder farmers in Kenya, Uganda and Ethiopia.* Radboud University Nijmegen: Centre for International Development Issues Nijmegen.

International Trade Centre (ITC) (2016). *Social and Environmental Standards: Contributing to More Sustainable Value Chains*. Geneva: ITC/EUI.

International Trade Centre (ITC). (2011). *The Impacts of Private Standards on Producers In Developing Countries.* (Literature Review Series on the Impacts of Private Standards; Part II). Geneva: International Trade Centre (ITC).

Jaffee, S., and Henson, S. (2004). *Standards and agro-food exports from developing countries: rebalancing the debate*. Washington, DC: World Bank.

John, J., and Mansingh, P. (2013). Plantation Labour Act 1951 and social cost – the case of tea industry. *NRPPD Discussion Paper No. 26*. National Research Programme on Plantation Development.

Johnston, Deborah and Abreu, Alexandre (2016). 'The asset debates: How (not) to use asset indices to measure well-being and the middle class in Africa.' *African Affairs*, 115. pp.399-418.

Kabeer, N. (2001). Conflicts over credit: re-evaluating the empowerment potential of loans to women in rural Bangladesh. *World Development*, 29(1), pp.63-84.

Kevane, M. (1994). 'Village labor markets in Sheikan district, Sudan.' *World Development*, 22(6), pp.839-857.

Kolk, A. (2005). Corporate social responsibility in the coffee sector: the dynamics of MNC responses and code development. *European Management Journal*, 23(2), pp.228–236.

Kumar, R., Nelson, V., Martin, A., Badal, D., Latheef, A., Reddy, B. S., Narayanan, L., Young, S. and Hartog, M. (2015). *Evaluation of the early impacts of the Better Cotton Initiative on smallholder cotton producers in Kurnool district India: baseline report*. Commissioned by ISEAL and the Ford Foundation, Natural Resources Institute, University of Greenwich report, Chatham.

Loconto, A., and Busch, L. (2010). Standards, techno-economic networks, and playing fields: Performing the global market economy. *Review of International Political Economy*, 17(3), pp.507-536.

Lipsey, M., and Wilson, D. (2001). *Practical meta-analysis*. London: Sage Publications.

Maertens, M., and Swinnen, J. F. (2009). Trade, standards, and poverty: evidence from Senegal. *World Development*, 37(1), pp.161-178.

Masood, A. (2014). *GlobalGAP certification and international trade flows*. PhD. University of Gottingen.

McCulloch, N., Winters, L. A., and Cirera, X. (2001). *Trade liberalization and poverty: a handbook.* London: UK Department for International Development and the Centre for Economic Policy Research.

Mezzadri, Alessandra (2014) Backshoring, local sweatshop regimes and CSR in India. *Competition and Change*, 18(4), pp.327-344.

Morrison, J. and Murphy, S. (2004). *Agricultural Trade and Poverty Reduction: Opportunity or Threat?* London: DFID.

Muradian, R., and Pelupessy, W. (2005). Governing the coffee chain: the role of voluntary regulatory systems. *World Development*, 33(12), pp.2029–2044.

Neilson, J., and Toth, R. (2016). *Baseline report: evaluation of the early impacts of sustainability standards on smallholder coffee farmers in Lampung and South Sumatra, Indonesia*. Sydney: University of Sydney.

Nelson, V. and Martin, A. (2011). *Impact Evaluation of Social and environmental Voluntary Standard Systems (SEVSS): Using theories of change*. University of Greenwich: Natural Resources Institute.

Nelson, V., and Pound, B. (2009). *The last ten years: a comprehensive review of the literature on the impact of Fairtrade*. London: Fairtrade Foundation.

Nicholls, A., and Opal, C. (2004). *Fair trade – market-driven ethical consumption*. London: Sage Publications.

Ouma, S. (2015). *Assembling export markets: the making and unmaking of global food connections in West Africa*. John Wiley & Sons.

Oya, C. (2012). Contract Farming in Sub-Saharan Africa: A Survey of Approaches, Debates and Issues. *Journal of Agrarian Change*, 12(1), pp.1-33.

Oya, C., & Pontara, N. (Eds.). (2015). *Rural Wage Employment in Developing Countries: Theory, Evidence, and Policy*. Routledge.

Oya, C., Johnston, D., Muchiri, E., Schaefer, F., Skalidou, D., Dickson, K., and Stansfield, C. (2015). *Protocol: Effects of certification systems for agricultural commodity production on socio-economic outcomes in low and middle-income countries: a systematic review*. London: The Campbell Collaboration.

Piza, C., Cravo, T., Taylor, L., Gonzalez, L., Musse, I., Furtado, I., Sierra, A. C., and Abdelnour, S. (2016). The impact of business support services for small and medium enterprises on firm performance in low- and middle-income countries: a systematic review. *Campbell Systematic Reviews*. 2016(1).

Ponte, S. (2009). Governing through quality: conventions and supply relations in the value chain for South African wine. *Sociologia ruralis*, 49(3), pp.236-257.

ProForest. (2005). *Developing a mechanism for palm oil traceability from plantation to end user*. Discussion paper for RT3. Oxford: ProForest.

Prowse, M. (2012). *Contract farming in developing countries: a review*. Agence Française de Développement.

Rainforest Alliance. (2013). *Sustainability Certification Dramatically Improves Smallhodler Farms' Access to Credit.* [online] Available at: http://www.rainforest-alliance.org/press-releases/citi-finance-study.

Rainforest Alliance. (2014). *Rainforest Alliance Certified Cocoa.* [online]. Accessed on January 10, 2017. Available at: http://www.rainforest-alliance.org/articles/rainforest-alliance-certified-cocoa

Randolph, J. J. and Edmondson, R. S. (2005). Using the binomial effect size display (BESD) to present the magnitude of effect sizes to the evaluation audience. *Practical assessment, research and evaluation*, 10(14), pp.1-7.

Ravallion, M. (2007). Evaluating anti-poverty programmes. *Handbook of Development Economics*, 4, pp.3788-3846.

Raynolds, L. T. (2000). Re-embedding global agriculture: the international organic and fair trade movements. *Agriculture and Human Values*, 17(3), pp.297-309.

Raynolds, L. T. (2009). Mainstreaming fair trade coffee: from partnership to traceability. *World Development*, 37(6), pp.1083-1093.

Raynolds, L. T., Murray, D., and Wilkinson, J. (Eds.). (2007). *Fair trade: the challenges of transforming globalization*. Routledge.

Raynolds, L. T., and Long, M. A. (2007). *Fair/alternative trade: historical and empirical dimensions* (pp. 15-32). In: Raynolds, L. T., Murray, D., & Wilkinson, J. (Eds.). *Fair trade: The challenges of transforming globalization*. Routledge.

Raynolds, L.T., and D. Murray. (2007). Fair Trade: contemporary challenges and future prospects. In Raynolds, L. T., Murray, D., & Wilkinson, J. (Eds.). (2007). *Fair trade: the challenges of transforming globalization*. Routledge. Pp. 223-234

Raynolds L.T., and N. Greenfield. (2015). Fair trade: movement and markets. In Raynolds L.T., and E.A. Bennett (ed.) *Handbook of research on fair trade*. Edward Elgar. Pp. 25-41

Reardon, T., Barrett, C.B., Berdegue, J.A., and Swinnen, J.F.M., (2009). 'Agrifood industry transformation and small farmers in developing countries'. *World Development* 37 (11), pp.1717–1727.

Riisgaard, L. (2009). How the market for standards shapes competition in the market for goods: sustainability standards in the cut flower industry, *DIIS Working Paper 2009:07*. Copenhagen: Danish Institute for International Studies (DIIS).

Ruben, R. (2013). *Critical review of some recently published Fair Trade impact studies: Deficient study design and little robust evidence*. Radboud University Nijmegen. Retrieved from: http://www.ru.nl/cidin/@883941/pagina/.

Ruben, R. (2012). The Fair Trade Balance: New Challenges after 25 years of Fair Trade. FERDI Policy Brief n. 52. http://www.ferdi.fr/en/publication/b52-fair-trade-balance-new-challenges-after-25-years-fair-trade

Saikia, B. (2008). Research notes and communications development of tea garden community and Adivasi identity politics in Assam. *Indian Journal of Labour Economics*, 51(2), pp.307-322.

Sandieson, R. (2006). Pathfinding in the research forest: the pearl harvesting method for effective information retrieval. *Education and Training in Developmental Disabilities*, 41(4), pp.401-409.

Sterne, J., Sutton, A., Ioannidis, J., Terrin, N., Jones, D., Lau, J., Carpenter, J., Rücker, G., Harbord, R., Schmid, C., Tetzlaff, J., Deeks, J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, G., Moher, D., and Higgins, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, 343:d4002.

Terstappen, V., Hanson, L., and McLaughlin, D. (2013). Gender, health, labor, and inequities: a review of the fair and alternative trade literature. *Agriculture and Human Values*, 30(1), pp.21-39.

Thomas, J., Brunton, J., and Graziosi, S. (2010). EPPI-Reviewer 4.0: software for research synthesis. EPPICentre software. London: Social Science Research Unit, Institute of Education.

Thomas, J., and Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(45).

Ton, G., Vellema, S., and Ge, L. (2014). The Triviality of Measuring Ultimate Outcomes: Acknowledging the Span of Direct Influence. *IDS Bulletin*, 45(6), pp.37-48.

Ton, G., Vellema, S., Desiere, S., Weituschat, S., and D'Haese, M. (2016). *Effectiveness of contract farming for improving income of smallholder farmers: preliminary results of a systematic review*. Paper prepared for: 149th EAAE Seminar. Rennes, France, October 27-28.

Tropical Commodity Coalition. (2012). *Coffee Barometer 2012*. teacoffeecocoa.org. Retrieved on April 28, 2013, from http://issuu.com/spanhuysen/docs/ttc_coffeebarometer.

USDA. (2006). *Tropical Products: World Markets and Trade*. Foreign Agricultural Service.

UTZ. (2014). *How can UTZ address poverty in cocoa farming?.* [online]. Accessed on Janury 10, 2017. Available at:https://www.utz.org/better-business-hub/strengthening-your-reputation/prosperity-for-cocoa-farmers-just-around-the-corner/

Vaessen, J., Rivas, A., Duvendack, M., Palmer Jones, R., Leeuw, F. L., Van Gils, G., Lukach, R., Holvoet, N., Bastiaensen, J., Hombrados, J. G., and Waddington, H. (2014). The effects of microcredit on women's control over household spending in developing countries: a systematic review and meta-analysis. *Campbell Systematic Reviews*, 2014(8).

Vagneron, I., and Roquigny, S. (2010). *What do we really know about the impact of fair trade? A synthesis*, Paris: PFCE.

Von Hagen, O., Manning, S., and Reinecke, J. (2010). Sustainable sourcing in the food industry: global challenges and practices. *Moderne Ernaehrung Heute*, Official Journal of the Food Chemistry Institute of the Association of the German Confectionery Industry, 4, pp.1-9.

Vorley, B *et al*. (2007) *Regoverning markets: a place for small-scale producers in the modern agrifood chains?* Aldershot: Gower Publishing.

Waddington, H. (2014) *Effect size calculation*. Presentation in 3ie training programme for systematic reviews, London, November.

Waddington, H., Snilstveit, B., Hombrados, J., Vojtkova, M., Phillips, D., Davies, P., and White, H. (2014) Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review. *Campbell Systematic Reviews*, 2014(6).

World Fair Trade Organization. (2017). *10 Principles of Fair Trade.* [online]. Accessed on January 10, 2017. Available at: http://wfto.com/fair-trade/10-principles-fair-trade

Winters, L. A. (2002). Trade liberalisation and poverty: what are the links? *World Economy*, 25, pp.1339–1367.

Other publications in the 3ie Systematic Review Series

The following reviews are available at
http://www.3ieimpact.org/publications/systematic-review-publications/

Short-term WASH interventions in emergency response: a systematic review. 3ie Systematic Review 33. Yates, T, Allen, J, Joseph, ML and Lantagne, D (2017)

Community monitoring interventions to curb corruption and increase access and quality of service delivery in low- and middle-income countries. 3ie Systematic Review 32. Molina E, Carella L, Pacheco A, Cruces, G and Gasparini, L (2016)

Effects and mechanisms of market-based reforms on access to electricity in developing countries: a systematic review. 3ie Systematic Review 31. Bensch, G, Sievert, M, Langbein, J, Kneppel, N (2016)

Youth gang violence and preventative measures in low- and middle-income countries: a systematic review (Part II), 3ie Systematic Review 30. Higginson, A, Benier, K, Shenderovich, Y, Bedford, L, Mazerolle, L, Murray, J (2016)

Youth gang membership and violence in low- and middle-income countries: a systematic review (Part I), 3ie Systematic Review 29. Higginson, A, Benier, K, Shenderovich, Y, Bedford, L, Mazerolle, L, Murray, J (2016)

Cash-based approaches in humanitarian emergencies: a systematic review, 3ie Systematic Review Report 28. Doocy, S and Tappis, H (2016)

Factors affecting uptake of voluntary and community-based health insurance schemes in low-and middle-income countries: a systematic review, 3ie Systematic Review 27. Panda, P, Dror, IH, Koehlmoos, TP, Hossain, SAS, John, D, Khan, JAM and Dror, DM (2016)

Parental, community and familial support interventions to improve children's literacy in developing countries: a systematic review, 3ie Systematic Review 26. Spier, E, Britto, P, Pigott, T, Roehlkapartain, E, McCarthy, M, Kidron, Y, Song, M, Scales, P, Wagner, D, Lane, J and Glover, J (2016)

Business support for small and medium enterprises in low- and middle-income countries: a systematic review, 3ie Systematic Review 25. Piza, C, Cravo, T, Taylor, L, Gonzalez, L, Musse, I, Furtado, I, Sierra, AC and Abdelnour, S (2016)

Interventions for improving learning outcomes and access to education in low- and middle- income countries: a systematic review, 3ie Systematic Review 24. Snilstveit, B, Stevenson, J, Phillips, D, Vojtkova, M, Gallagher, E, Schmidt, T, Jobse, H, Geelen, M, Pastorello, M, and Eyers, J (2015)

Economic self-help group programmes for improving women's empowerment: a systematic review, 3ie Systematic Review 23. Brody, C, De Hoop, T, Vojtkova, M, Warnock, R, Dunbar, M, Murthy, P and Dworkin, SL (2016)

The identification and measurement of health-related spillovers in impact evaluations: a systematic review, *3ie Systematic Review 22*. Benjamin-Chung, J, Abedin, J, Berger, D, Clark, A, Falcao, L, Jimenez, V, Konagaya, E, Tran, D, Arnold, B, Hubbard, A, Luby, S, Miguel, E and Colford, J (2015)

The effects of school-based decision-making on educational outcomes in low- and middle-income countries: a systematic review, 3ie Systematic Review Report 21. Carr-Hill, R, Rolleston, C, Pherali, T and Schendel, R, with Peart, E, and Jones, E (2015)

Policing interventions for targeting interpersonal violence in developing countries: a systematic review, *3ie Systematic Review 20*. Higginson, A, Mazerolle, L, Sydes, M, Davis, J, and Mengersen, K (2015)

The effects of training, innovation and new technology on African smallholder farmers' wealth and food security: a systematic review, *3ie Systematic Review 19*. Stewart, R, Langer, L, Rebelo Da Silva N, Muchiri, E, Zaranyika, H, Erasmus, Y, Randall, N, Rafferty, S, Korth, M, Madinga, N and de Wet, T (2015)

Community based rehabilitation for people with disabilities in low- and middle-income countries: a systematic review, *3ie Systematic Review 18*. Iemmi, V, Gibson, L, Blanchet, K, Kumar, KS, Rath, S, Hartley, S, Murthy, GVS, Patel, V, Weber, J and Kuper H (2015)

Payment for environmental services for reducing deforestation and poverty in low- and middle-income countries: a systematic review, *3ie Systematic Review 17.* Samii, C, Lisiecki, M, Kulkarni, P, Paler, L and Chavis, L (2015)

Decentralised forest management for reducing deforestation and poverty in low- and middle- income countries: a systematic review, *3ie Systematic Review 16.* Samii, C, Lisiecki, M, Kulkarni, P, Paler, L and Chavis, L (2015)

Supplementary feeding for improving the health of disadvantaged infants and young children: a systematic and realist review, *3ie Systematic Review 15.* Kristjansson, E, Francis, D, Liberato, S, Greenhalgh, T, Welch, V, Jandu, MB, Batal, M, Rader, T, Noonan, E, Janzen, L, Shea, B, Wells, GA and Petticrew, M (2015)

The impact of land property rights interventions on investment and agricultural productivity in developing countries: a systematic review, 3ie Systematic Review Report 14. Lawry, S, Samii, C, Hall, R, Leopold, A, Hornby, D and Mtero, F, 2014.

Slum upgrading strategies and their effects on health and socio-economic outcomes: a systematic review, 3ie Systematic Review 13. Turley, R, Saith, R., Bhan, N, Rehfuess, E, and Carter, B (2013)

Services for street-connected children and young people in low- and middle-income countries: a thematic synthesis, 3ie Systematic Review 12. Coren, E, Hossain, R, Ramsbotham, K, Martin, AJ and Pardo, JP (2014)

Why targeting matters: examining the relationship between selection, participation and outcomes in farmer field school programmes, 3ie Systematic Review 11. Phillips, D, Waddington, H and White, H (2015)

The impact of export processing zones on employment, wages and labour conditions in developing countries, 3ie Systematic Review 10. Cirera, X and Lakshman, R (2014)

Interventions to reduce the prevalence of female genital mutilation/cutting in African countries, 3ie Systematic Review 9. Berg, RC and Denision, E (2013)

Behaviour change interventions to prevent HIV among women living in low and middle income countries, 3ie Systematic Review 8. McCoy, S, Kangwende, RA and Padian, NS (2009)

The impact of daycare programs on child health, nutrition and development in developing countries, 3ie Systematic Review 7. Leroy, JL, Gadsden, P and Guijarro, M (2011)

Willingness to pay for cleaner water in less developed countries: Systematic review of experimental evidence, 3ie Systematic Review 6. Null, C, Hombrados, JG, Kremer, M, Meeks, R, Miguel, E and Zwane, AP (2012)

Community-based intervention packages for reducing maternal morbidity and mortality and improving neonatal outcomes, 3ie Systematic Review 5. Lassi, ZS, Haider, BA and Langou, GD (2011)

The effects of microcredit on women's control over household spending: a systematic review, 3ie Systematic Review 4. Vaessen, J, Rivas, A, Duvendack, M, Jones, RP, Leeuw, F, van Gils, G, Lukach, R, Holvoet, N, Bastiaensen, J, Hombrados, JG and Waddington, H, (2013).

Interventions in developing nations for improving primary and secondary school enrolment of children: a systematic review, 3ie Systematic Review 3. Petrosino, A, Morgan, C, Fronius, T, Tanner-Smith, E, and Boruch, R, 2016.

Interventions to promote social cohesion in Sub-Saharan Africa, 3ie Systematic Review 2. King, E, Samii, C and Snilstveit, B (2010)

Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries, 3ie Systematic Review 1. Waddington, H, Snilstveit, B, White, H and Fewtrell, L (2009)

International Initiative for Impact Evaluation
London International Development Centre
36 Gordon Square
London WC1H 0PD
United Kingdom

3ieuk@3ieimpact.org
Tel: +44 207 958 8351/8350

www.3ieimpact.org