

Prof. Dr. Stefan K. Lhachimi, MPP, MSc

Research Group for Evidence-Based Public Health, University of Bremen,
Germany: www.ebph.uni-bremen.de

Call Replication Window 4
Technical Proposal for:
Cash, food, or vouchers?
Evidence from a randomized experiment in northern Ecuador¹
by
Melissa Hidrobo, John Hoddinott, Amber Peterman,
Amy Margolies, and Vanessa Moreira
June 2017

Contents

1	Overview.....	2
2	Replication Plan.....	5
2.1	Push Button Replication.....	5
2.2	Pure Replication.....	5
2.3	Methodology and Estimation Analysis (MEA).....	6
2.3.1	Sampling Frame.....	6
2.3.2	Generalized Linear Mixed Models as competing modeling approaches.....	7
2.3.3	Cost data.....	8
2.4	Theory of Change Analysis (TCA).....	8
3	Acknowledgments.....	14
4	References.....	14

¹ in *Journal of Development Economics*, Volume 107, March 2014, p 144-156.

1 Overview

The original paper by Hidrobo et al. (hereafter referred to as: the paper) analyses a cluster-randomized controlled trial (cRCT) comparing the effectiveness of different modes of food assistance (cash, food, and vouchers) with a control mode (i.e. no assistance). The statistical analysis was done using an ANCOVA model and included several food-related outcome measures (food consumption, several indices of food security, and diet). Robustness checks were made and the effect estimates, adjusted for co-variates, are reported. Finally, the costs associated with each mode of assistance were calculated and the cost-effectiveness measures presented.

Food assistance in order to counter mal- or undernutrition among the vulnerable parts of the population is an ongoing concern for many countries, most notably for low and middle-income countries where social protection mechanisms are often under developed. Mal- or undernutrition can have severe long-term consequence on human capital, in particular on the ability to study and work (Marmot et al., 2012). General food subsidies, e.g. fixing the price of certain food commodities, have been shown to be highly inefficient in targeting the neediest part of the population while being prohibitively expensive for many governments in the long run. Recently, different types of social assistance interventions targeting the most vulnerable part of the population have become more common and have gained prominence on development agendas in what is often considered a "quiet revolution"(Barrientos and Hulme). Social assistance interventions are usually defined as "noncontributory transfer programs targeted in some manner to the poor and those vulnerable to poverty and shocks" (World Bank, 2011) to ensure an adequate standard of living and ensure long-term health. Social assistance interventions are often differentiated into cash transfers; in-kind transfers; fee waivers/vouchers; subsidies; and public works programs (Pega et al., 2015). All three modes investigated by the paper (cash, in-kind, vouchers) are under policy discussion or even used in several countries. However, recent systematic reviews show that evidence based on high-quality

studies, i.e. randomized controlled trials that conduct a head-to-head comparison between these modes, is exceedingly rare (Pega et al., 2015, Pega et al., 2017 (forthcoming), Lagarde et al., 2009). Many studies are either observational or do compare one of the three interventions against a non-intervention only. However, comparing these three modes directly against each other is of vital importance for policy makers as these three modes are thought to have distinct advantages/disadvantages in terms of efficacy, public acceptance, and cost. For example, direct cash transfers are seen by some as a more efficient way of providing help as the disbursement costs are relatively low and allow the recipient to buy goods that truly increase the recipient's utility (Fiszbein and Schady, 2009). On the other hand, recipients may spend cash not necessarily solely on beneficial goods (e.g. for tobacco instead). In-kind transfers, for example, may have more beneficial health effects if the quality and quantity of food provided exceeds that bought from a cash transfer. Yet in-kind transfers are costly to administer and reduce the agency of the recipients. Hence, some argue that vouchers occupy a middle ground between these two modes of assistance.

The paper by Hidrobo et al is one of the very few studies that report the results of a head-to-head comparison of all three modes. Although the authors clearly described the analysis, the underlying study may have some limitations. For example, although the study clearly demonstrates that all three modes of food assistance are superior to the control situation, Hidrobo and colleagues do not give a clear recommendation as to which mode of food assistance is superior, e.g. by pointing out that such a recommendation mainly depends on the preference of the policymaker. This is a fact and, whereas cost-effectiveness is not (and should not) be the only criterion for a policy decisions, a full cost-effectiveness analysis at the outcome level is usually considered to be helpful for the decision makers. If the primary aim of their study was to establish which of the three modes of food assistance is superior (in terms of food security, costs, and acceptance), the sampling frame and size of the study appear to be underpowered to identify a meaningful difference between the three modes (although

the study is sufficiently powered to reveal a difference between the control mode and the interventions). In the paper, the sampling frame is not well described (probably due to restrictions related to article length) and requires further scrutiny. In addition, although the statistical model used is appropriate and parsimonious, a competing statistical model, i.e. a generalized linear mixed model, may be more fitting considering the sampling frame, cluster distribution, and the number of variables.

Lastly, the analysis of the costs and the derived cost-effectiveness analysis is underdeveloped. Ideally, uncertainty in costs and effects should be modeled jointly. However, as this may prove difficult (because costs were not empirically measured but derived from accounting data) a simulation analysis might provide additional insight. This could then inform policymakers about the potential effect of uncertainty and variability in the cost items on the cost-effectiveness for several endpoints. Moreover, this could help to design future studies by considering what sample size is required to find a statistically significant difference between these different modes.

2 Replication Plan

This final version of the replication plan incorporates comments from: i) several reviewers, ii) the external adviser, and iii) the authors of the original paper (Hidrobo et al.). When drafting this version, the replication team was already in possession of the data and the replication code, but had not yet interacted with the material provided.

The push button replication (PBR) is the first step of interacting with the data as provided by the authors. This step will be documented with a dedicated final memorandum and reported as a separate section of the final replication report. In a second step, a pure replication will be conducted based on the original data; here, the replication team aims to reproduce the results of the original paper. Further, a cost-effectiveness analysis is performed to add further value (as mentioned above) as part of the theory of change analysis.

2.1 Push Button Replication

The authors used the software package **Stata**. Hidrobo et al. have already provided access to the relevant data and the replication code.

2.2 Pure Replication

For the pure replication, we will reproduce the analysis in **R** and (potentially) in **WinBugs**. **R** is particularly suited for dynamic documentation of the analysis (e.g. through *knitr*). Thus, the replication will be available with two different software packages, one of which will be ‘open source’ and the other ‘commercial’. The script for the open access version will be fully commented and (depending on approval, privacy protection, and copyright issues) the script can be made publicly available on a dedicated server (e.g. www.ebph.uni-bremen.de). The pure replication will be carried out in **R** with the aim of reproducing the findings of the original study based on the data received and the information provided in the original paper.

2.3 Methodology and Estimation Analysis (MEA)

The original analysis calculated unadjusted and adjusted treatment effects using ANCOVA. Robust standard errors were calculated at the cluster level (accounting for intra-cluster correlation, albeit the precise estimator was not given). However, it is unclear whether and how the sampling weights were included. The description of the sampling frame is not entirely sufficient and most of the suggestions proposed in this section aim to explicitly account for the hierarchical and stratified sampling frame of this cRCT.

2.3.1 Sampling Frame

The authors use a cRCT design and report all data by intervention arm (Chow and Liu, 2004). The authors state that randomization into control and intervention was done at the barrio level. Each barrio contains 1 to 6 clusters and these clusters are geographical units within the barrios. However, it is unclear whether the size and number of the clusters across the arms is adequately balanced. This is particularly important as it appears that, within one barrio, several clusters with different types of treatment interventions may exist (i.e. geographical separation was only assured for control vs. treatments, and not within the treatment arms). This may have resulted in some form of contamination between the treatments arms (Torgerson, 2001); however, identifying contamination and statistically correcting for this can be difficult and is often not possible (Keogh-Brown et al., 2007). Nevertheless, statistically modeling whether this structure has an influence on the outcomes may prove valuable for better understanding of the effect of the sampling frame, particularly as the number and (presumably) type of treatment cluster varies in each barrio. This can be done in a sensitivity analysis where the barrios with (likely) contamination are excluded from the model (in the re-analysis this could be done by a stepwise leave-one-out analysis to measure the sensitivity of the estimates). Another approach is to see whether correlation of the effects across treatment arms vary by the

number of clusters. Finally, no additional covariates at the cluster level seem to be included, e.g. population size, socio-economic structure, etc.²

2.3.2 Generalized Linear Mixed Models as competing modeling approaches

Generalized linear mixed models (GLMM) account for the hierarchical nature of the data (Gelman and Hill, 2007), e.g. barrios and clusters. This model class is also known as a hierarchical or multilevel model, and can address typical violations of assumptions of standard linear models, such as normality, homogeneity of variance, or independence of data. The latter is particularly important in the present application, as the data are hierarchical and the authors speak only of the “...relative success of random assignment” without giving further details, e.g. was randomization only successful at the treatment arm level and not at the cluster level? GLMM allows for the inclusion of fixed and random effects. Random effects take into account the non-independence of the observations at the cluster level. If the GLMM estimates for the intervention effects differ significantly from the originally estimated effect, we can conclude that the results are sensitive to the modeling choice. However, a significant difference may not necessarily translate into a substantial or ‘practically meaningful’ difference. This has to be judged on a qualitative basis, e.g. does the interpretation of the results change. Moreover, the GLMM analysis will yield the intra-class correlation between the clusters (Eldridge and Kerry, 2012). This statistic should be reported according to the CONSORT guidelines to allow proper inclusion of the estimates into a meta-analysis (Schulz et al., 2010). Additionally, we will also run a parametric bootstrap to calculate confidence interval compare those with standard GLMM results; parametric bootstrap is more robust to deviations of the underlying modeling assumptions (Bates et al., 2015).

² Hidrobo et al. pointed out that the Editor of the journal explicitly advised against the inclusion of additional covariates (personal communication with the authors).

2.3.3 Cost data³

In the paper, calculation of the cost data used is not extensively described (probably due to limitations of article size). Moreover, the cost data were not measured at the unit of observation, but were derived using data on global input and then calculated based on the assumption of equality across observations and without accounting for uncertainty. For example, the apparently substantial costs of negotiating with supermarkets may vary significantly between clusters or regions (Evans and Popova, 2016). Moreover, in the description of the cost data, it is unclear whether the cost per beneficiary takes the household composition into account (e.g. household size, or age structure) and whether these vary across clusters. Ideally, cost data could be re-calculated at the unit of observation; however, the authors informed us that such data are not available. Nevertheless, the apparent uncertainty surrounding the cost data calls for additional analysis that quantifies the extent of uncertainty and provides more information about how uncertainty could affect the cost-effectiveness of the different modes.

2.4 Theory of Change Analysis (TCA)

As outlined above, the cost analysis of the study is limited, since the presented cost data do not take into account the variability and uncertainty inherent to (cost) data. Moreover, from a health economic viewpoint⁴ the most appropriate measure of cost-effectiveness is the incremental cost-effectiveness ratio (ICER), defined as $\frac{C_1 - C_0}{E_1 - E_0}$, where the marginal costs of two interventions is divided by the marginal effect. The ICER shows the relative cost-effectiveness of an intervention to the 'next best', i.e. cost-

³ We consider the cost data as reported by the authors as part of the section *Methodology and Estimation Analysis*. However, additional steps will be made to analyze the cost data as part of the *Theory of Change Analysis* (see below).

⁴ For the standard health economic concepts mentioned below (e.g., *incremental cost effectiveness ratio, efficiency frontier, probabilistic sensitivity analysis*) we follow the standard examples found in widely accepted textbooks on health economic evaluation, such as DRUMMOND, M. 2015. *Methods for the economic evaluation of health care programmes*, Oxford, United Kingdom ; New York, NY, USA, Oxford University Press.

effective, intervention. The ICER has several advantages compared with using average cost-effectiveness ratios. In particular, when several interventions are compared simultaneously, the ICER allows to construct an efficiency frontier⁵ to maximize (health) effects given a budget constraint.

We propose to undertake a certain type of health economic simulation analysis, i.e. a probabilistic sensitivity analysis (PSA), for all three mode of food assistance to quantify the potential influence of uncertainty in the measurement of data or expected variability (e.g. market price fluctuations, or time spent on obtaining food).⁶ A PSA is often used to account for *uncertainty in the parameters* in health economic evaluations (sometimes also called second-order uncertainty).⁷

We will use a stylized, intuitive regression analogy to explain what is meant by second-order uncertainty in contrast to variability and heterogeneity (we follow the example of (Briggs et al., 2006). Think of a standard linear regression:

$$Y = \alpha + \sum_{i=1}^k \beta_i X_i + \epsilon$$

In this analogy, the dependent variable Y would be the output of the simulation model. The coefficients α and β are the input parameters of the simulation model. Now, the coefficient β models the *heterogeneity* in the sense that different values of X lead to different values of Y . To capture additional heterogeneity an additional characteristic (variable X_k) and an input parameter (β_k) are needed. The *parameter uncertainty* in a simulation model is similar to the standard errors of the regression estimates for the coefficients α and β_i . The *variability* of a simulation is represented by the error term *epsilon*. The *model uncertainty* refers to the question as to whether we can truly assume,

⁵ The efficiency frontier, similar to a production frontier, shows the set of the most efficient points in a cost-effectiveness plane as span by the most cost-effective interventions.

⁶ Please note, such a simulation generally cannot draw fundamentally new conclusions about how cost data have been collected. However, by reasonably varying the key inputs into this model it will show how sensitive the results are to variations in input.

⁷ A PSA could also be used to account for uncertainty in the model structure.

say, an additive regression model, or whether, say, a multiplicative specification would be more appropriate.

A PSA is run for several hundred or thousand times while, for each run, the input parameters are drawn from a stochastic distribution. For the purpose of our replication study, the effect parameters can be taken from the statistical estimation model, i.e., the effect estimates documented in the original paper and, if substantially different effect estimates are derived, from the GLMM model (see above). However, in this PSA exercise, the challenge lies in the lack of an estimator for the uncertainty of the cost data. We will employ several different approaches to derive an estimate for the cost data.⁸ However, none of these can substitute for the fact that these estimates are lacking. First, we will assume a standard error of 30% of the mean of the input parameter, i.e. the cost estimate. This is an often used initial and conservative assumption when standard errors are not available (Weinstein et al., Gray and Clarke, 2011). Second, we will review cost estimates and the reported variability from similar programs (e.g. conditional cash transfers in South American countries compared to in-kind transfers).⁹ These estimates may provide additional understanding about the levels of uncertainty that can reasonably be expected in such data (please note: we do not use the point estimates reported, but the magnitude of the uncertainty). In both cases we will use widely agreed distributional assumptions for cost data as inputs (e.g. log-normal or gamma distribution) (Briggs et al., 2012, Briggs et al., 2006). Another challenge is that the input parameters of a PSA are correlated with each other. That is, treatment effects across the intervention arms are, in general, positively correlated (e.g. because of common but unobservable causal factors). Similarly, costs are, in general, also positively correlated. However, these correlations can only be estimated when data at the individual level are available.

⁸ Initially, we hoped that cost data had been collected at the individual level but were not reported in the original paper; however, the authors informed us that this is not the case.

⁹ This will be mainly based on the studies included in the systematic review of Lagarde et al. 2009 and of Pega et al. 2017.

Hence, as a robustness check of our PSA, we will run a scenario where we assume a correlation among treatment as used in a similar exercise in which empirical correlations were not available (Stollenwerk et al., 2015).

A full PSA will yield several thousand ICERS, i.e. one estimate for each draw from the stochastic distribution of the input parameters.¹⁰ These will then be depicted in a cost-effectiveness plane, where the x-axis measures the incremental change in effects and the y-axis the incremental costs (see Figure 1 for a stylized numerical example). An individual ICER (as calculated by an individual run of a full PSA) will lie in one of the four quadrants of the cost-effectiveness plan. An ICER lying in the north-west quadrant shows that the new intervention is clearly dominated by the comparator, i.e. less effective and more costly, while in the SE quadrant the opposite is true, i.e. more effective and less costly. The most interesting quadrant is the NE quadrant where the decision as to whether an intervention is considered cost-effective depends on the willingness-to-pay (WTP) of the decision-maker. Due to the stochastic nature of the simulation, estimates of ICERs will often be found in all four quadrants. The distribution of the ICERs across the plane will then give information about the variability in terms of cost-effectiveness between the interventions, and enable us to understand to what extent the cost-effectiveness of the interventions depends on the variability of the included costs.

After a full PSA, we will then calculate the percentage of cases for which an intervention is more cost-effective for a given WTP, and estimate this number for varying WTP thresholds.¹¹ Calculating this

¹⁰ Because the effect estimate will be done at the outcome level, several CE planes will be reported.

¹¹ Calculation of a confidence interval around an ICER estimate is mathematically very difficult and often impossible, as an ICER is the ratio of two stochastic distributions. This ratio may not be tractable.

metric will provide information as to which intervention has the highest likelihood to be considered cost-effective. For policymakers, extending the cost-effectiveness analysis of the original paper in this way will provide important information on the comparative cost-effectiveness of the examined interventions in the original study. In the worst case, no difference will be detected between the three modes, even when assuming only minor uncertainty in the cost data. This could show that, irrespective of the underlying quality of the cost data, the interventions do not differ from each other in terms of their cost-effectiveness, so that decisions about the use of a certain mode should be made without overly relying on the difference in costs. In the best case, the difference in cost-effectiveness between modes will be robust, even with major uncertainty in the cost data. Policymakers can then be more certain as to which mode is the most cost-effective intervention. The most likely scenario is a middle ground that (by varying the magnitude of uncertainty in the cost data) the cost-effectiveness of the modes cannot be clearly ranked. However, this information can be used to inform and justify a new study design, e.g. what sample size is required to robustly investigate differences in the cost-effectiveness between the modes in a new trial.

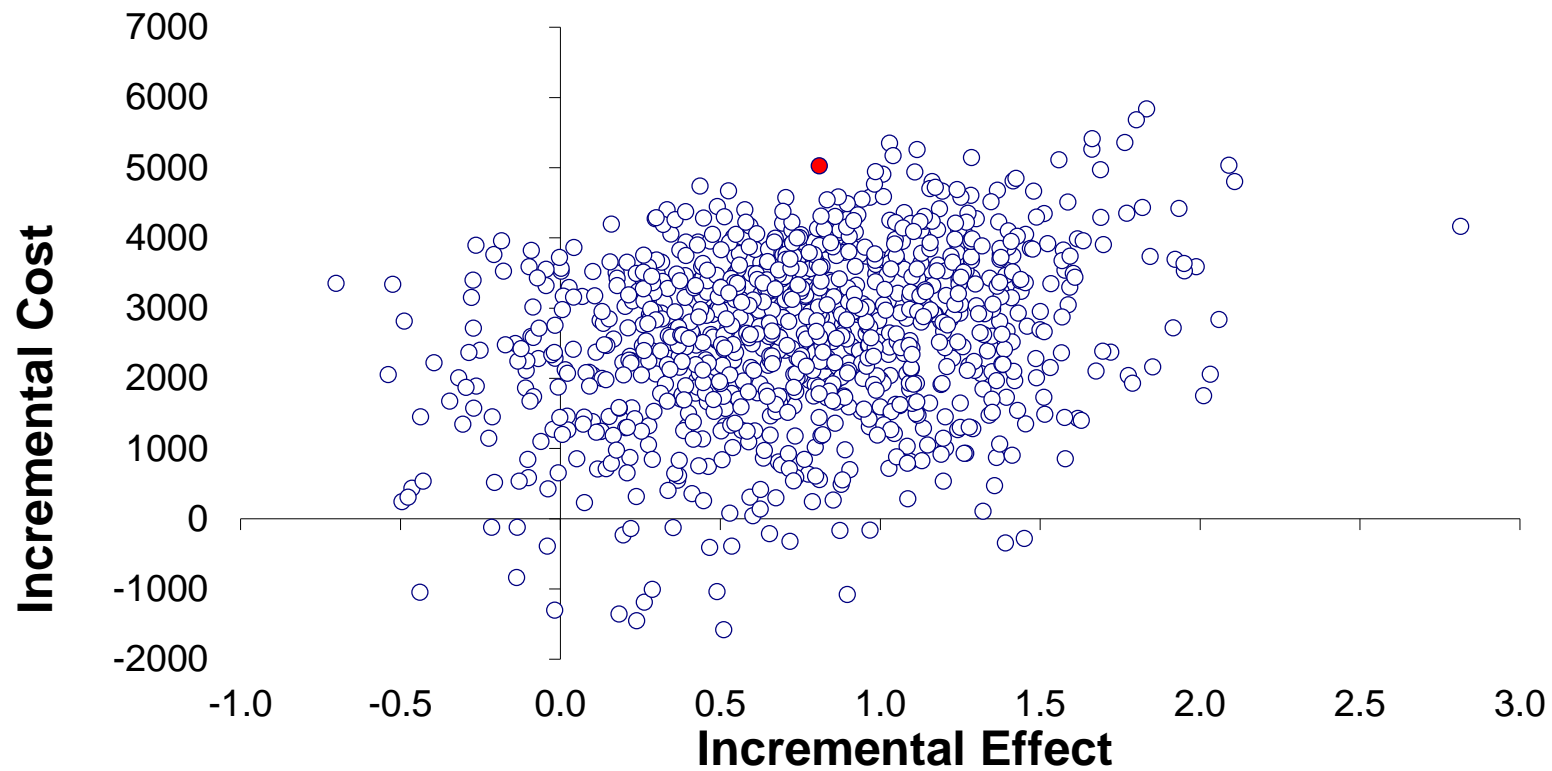


Figure 1: Stylized Cost-Effectiveness Plane.

The solid dot in the north-east quadrant is the incremental cost-effectiveness ratio (ICER) of the i -th run of a probabilistic sensitivity analysis. The point depicts that the new intervention has an incremental effect of 0.8 and an incremental cost of 5,000 as compared to the old intervention. Whether this can be considered cost-effective relies on the willingness-to pay, i.e. if the willingness to pay for 1 additional unit of effects is less than 6,250, then the new intervention (given the input data that has been used for the realization of the i -th ICER) is not cost-effective. For example, In the south-east quadrant, all the ICERs simulated with the respective input data are cost-effective (all have a positive incremental effect and negative incremental costs compared to the old intervention).

3 Acknowledgments

Thanks go to our collaborator Till Seuring (BIPS, Germany), the three anonymous reviewers, the participants of the 3ie workshop in Washington (D.C.), and to the authors of the original paper for their valuable comments.

4 References

- BARRIENTOS, A. & HULME, D. Social protection for the poor and poorest in developing countries: Reflections on a quiet revolution. *BWPI Working Paper 30*. Manchester, United Kingdom: University of Manchester.
- BATES, D., MÄCHLER, M., BOLKER, B. & WALKER, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *2015*, 67, 48.
- BRIGGS, A. H., CLAXTON, K. & SCULPHER, M. J. 2006. *Decision modelling for health economic evaluation*, Oxford, Oxford University Press.
- BRIGGS, A. H., WEINSTEIN, M. C., FENWICK, E. A., KARNON, J., SCULPHER, M. J., PALTIEL, A. D. & FORCE, I.-S. M. G. R. P. T. 2012. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*, 32, 722-32.
- CHOW, S.-C. & LIU, J.-P. 2004. *Design and analysis of clinical trials : concept and methodologies*, Hoboken, N.J., Wiley-Interscience.
- DRUMMOND, M. 2015. *Methods for the economic evaluation of health care programmes*, Oxford, United Kingdom ; New York, NY, USA, Oxford University Press.
- ELDRIDGE, S. & KERRY, S. 2012. *A practical guide to cluster randomised trials in health services research*, Chichester, Wiley.
- EVANS, D. K. & POPOVA, A. 2016. Cost-Effectiveness Analysis in Development: Accounting for Local Costs and Noisy Impacts. *World Development*, 77, 262-276.
- FISZBEIN, A. & SCHADY, N. 2009. *Conditional cash transfers: Reducing present and future poverty*, Washington, DC, World Bank.
- GELMAN, A. & HILL, J. 2007. *Data analysis using regression and multilevel/hierarchical models*, New York, Cambridge University Press.
- GRAY, A. & CLARKE, P. M. 2011. *Applied methods of cost-effectiveness analysis in health care*, Oxford ; New York, Oxford University Press.
- KEOGH-BROWN, M. R., BACHMANN, M. O., SHEPSTONE, L., HEWITT, C., HOWE, A., RAMSAY, C. R., SONG, F., MILES, J. N., TORGERSON, D. J., MILES, S., ELBOURNE, D., HARVEY, I. & CAMPBELL, M. J. 2007. Contamination in trials of educational interventions. *Health Technol Assess*, 11, iii, ix-107.
- LAGARDE, M., HAINES, A. & PALMER, N. 2009. The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries. *Cochrane Database Syst Rev*, CD008137.
- MARMOT, M., ALLEN, J., BELL, R., BLOOMER, E., GOLDBLATT, P., CONSORTIUM FOR THE EUROPEAN REVIEW OF SOCIAL DETERMINANTS OF, H. & THE HEALTH, D. 2012. WHO European review of social determinants of health and the health divide. *Lancet*, 380, 1011-29.

- PEGA, F., WALTER, S., LIU, S. Y. & LHACHIMI, S. K. 2015. Unconditional cash transfers for assistance in humanitarian disasters: effect on use of health services and health outcomes in low-and middle-income countries. *Cochrane Database of Systematic Reviews* [Online].
- PEGA, F., WALTER, S., LIU, S. Y., PABAYO, R., SAITH, R. & LHACHIMI, S. K. 2017 (forthcoming). Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries. *The Cochrane Library*.
- SCHULZ, K. F., ALTMAN, D. G., MOHER, D. & GROUP, C. 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med*, 7, e1000251.
- STOLLENWERK, B., LHACHIMI, S. K., BRIGGS, A., FENWICK, E., CARO, J. J., SIEBERT, U., DANNER, M. & GERBER-GROTE, A. 2015. Communicating the parameter uncertainty in the IQWiG efficiency frontier to decision-makers. *Health Economics*, 24, 481–490.
- TORGERSON, D. J. 2001. Contamination in trials: is cluster randomisation the answer? *BMJ*, 322, 355.
- WEINSTEIN, M. C., O'BRIEN, B., HORNBERGER, J., JACKSON, J., JOHANNESON, M., MCCABE, C. & LUCE, B. R. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices & Modeling Studies. *Value in Health*, 6, 9-17.
- WORLD BANK. 2011. *Safety Nets and Transfers: Glossary* [Online]. Available: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTSAFETYNETSANDTRANSFERS/0,,contentMDK:22200792~pagePK:210058~piPK:210062~theSitePK:282761,00.html> [15. July 2017] [Accessed].