# 3ie Impact evaluation screening protocol

**July 2019**

## 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, how, why and at what cost. We believe that using better and policy-relevant evidence helps to make development more effective and improve people's lives.

## 3ie Impact Evaluation Repository

The Impact Evaluation Repository (IER) is the largest database of published impact evaluation studies of development interventions in low-and middle-income countries. It includes impact evaluations published in English, Spanish and Portuguese. 3ie populates the repository through a systematic search and screening process. Over 35 databases, search engines, and websites are searched in order to locate all published development impact evaluations. Studies found through the search are then screened to ensure they meet the inclusion criteria. 3ie regularly updates the IER. For each update, staff review the existing protocol and revise it to take into account improvements to the updating criteria and process. We have published updates in 2014 and 2016.

## About the screening tool

3ie uses this screening tool to determine whether studies are eligible for inclusion in the IER. Among other things, it identifies the minimum methodological requirements for studies to be included. The methodological criteria deliberately strike a balance between being overly stringent or overly lenient. Inclusion in the IER does not indicate that a study is as rigorous as it could possibly be. Rather, it indicates that the study makes a credible claim to identifying the causal impact of an intervention. Our goal is to assemble studies that meet basic criteria for causal inference, while allowing users to apply their own standards in interpreting the results of particular studies they find in the IER.

This screening tool is one component of the overall search and screening protocol for the IER, which includes a comprehensive search strategy tailored to each database, a multi-stage screening process, and a data extraction process to classify studies by theme, sector, region, attention to equity and/or gendered inequality, and other features. Protocols from previous updates are available on the 3ie website.

**Suggested citation:** International Initiative for Impact Evaluation (3ie), 2019. *Impact Evaluation Repository screening tool.* New Delhi: International Initiative for Impact Evaluation (3ie).

# 3ie Impact evaluation screening protocol

| Instructions | | | | |
|---|---|---|---|---|
| Proceed through the questions in order. Note that an "unclear" answer never excludes a study. The questions are designed to be as objective as possible. The questions are meant to start with those easier to ascertain and progress to those that will be harder to answer based on a quick read. The screener should feel confident of any "yes" or "no" answer used to exclude a study. **If you cannot conclusively say "yes" or "no", please mark the study as unclear and it will move on to the next level of screening.** | | | | |
| | | **No** | **Yes** | **Unclear** |
| **1** | Is the publication date 1990 or after? | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **2** | Does the study concern a population within a country or countries classified as low- or middle-income?<br><br>***Note: See Appendix B for classifications.*** | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **3** | Does the study concern a policy, programme or intervention? | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **4** | Is the study a biomedical (efficacy) trial of a product, medication, or procedure? These include medical technologies.<br><br>*Typically, efficacy studies examine treatment outcomes under highly controlled conditions. Effectiveness studies go beyond laboratory trials and examine interventions in real world settings. Note that randomised controlled trials (RCTs) that only address the biomedical efficacy of a drug or treatment should be excluded.*<br><br>***Note: See Appendix C for further guidelines*** | | | |
| | **IF YES, THEN EXCLUDE** | | | |
| **5** | Does the study have a sample size of at least 50 observations for experimental and at least 100 observations for quasi-experimental methods at baseline (control and treatment combined)?<br><br>*Note: Cluster randomised evaluations must randomise at least **four** clusters*<br><br>*Note: we are in the process of revising the minimum sample size criteria, so that they no longer specify exact numbers and instead reflect the fact that the minimum sample size required for a rigorous evaluation depends on the study design. If you have doubts about whether a study's sample size is adequate, you may flag it for further review.* | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **6** | Is the study described as a systematic review, synthetic review, and/or meta-analysis? | | | |
| | **IF YES, THEN EXCLUDE** | | | |
| **7** | Is the study completed?<br><br>*Note: Published protocols or baseline findings, impact evaluation designs, and process evaluations are not included in the IER.* | | | |
| | **IF NO, THEN EXCLUDE** | | | |

| | | | | |
|---|---|---|---|---|
| **8** | Are impact evaluation results reported?<br><br>*Note: Published protocols or baseline findings, and process evaluations are not included in the IER.*<br><br>*Note: The IER does not accept cost-benefit or cost-effectiveness analyses that **do not** report outcomes for an impact evaluation.* | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **9** | Does the study evaluate the **effectiveness** of a policy, programme or intervention?<br><br>*Note: Feasibility or acceptability studies are not accepted.* | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **10** | Is the study's primary identification or estimation strategy one or any of the following:<br><br>a) Randomised evaluation (includes RCTs, cluster RCTs, social experiments, random assigned studies, randomised field trials or randomised controlled experiments)<br>b) Propensity score matching (PSM) or other matching methods (as well as synthetic controls)<br>c) Difference-in-differences (DID), or a fixed or random effects model with an interaction term between time and intervention for baseline and follow-up observation<br>d) Instrumental variable (IV) estimation (or other methods using an instrumental variable such as the Heckman Two Step approach)<br>e) Regression discontinuity design (RDD) or fuzzy-RDD.<br>f) Interrupted time series (ITS)<br>*Note: for ITS studies, the minimum reporting requirements are under development. For now, if an ITS study meets all other inclusion criteria, flag it for further review rather than making an include or exclude decision.*<br><br>*Note: The study may also use methods in addition to those listed here (such as regression with controls), or may use a primary evaluation methodology not listed (such as in a natural experiment), but must do so **in addition** to one of the above methods (a-e).*<br><br>***Note: See Appendix A for further guidelines.*** | | | |
| | **IF NO, THEN EXCLUDE** | | | |
| **11** | Is the study published in a journal, as a working paper or as an institutional report?<br><br>*Note: The IER does not accept ongoing studies (published protocols or designs), published drafts with no institutional affiliation or theses.*<br>*Note: Technical reports/paper and discussion papers are included as working papers if they are part of a series* | | | |
| | **IF NO, THEN EXCLUDE** | | | |

| | | | | |
|---|---|---|---|---|
| **12** | Does the intervention focus on any of the following:<br><br>a) prevention and treatment of non-communicable diseases, including cancer, cardiovascular diseases (e.g. heart attacks and stroke), chronic respiratory diseases (e.g. chronic obstructive pulmonary disease and asthma), arthritis and diabetes<br><br>b) prevention and treatment of mental illnesses, substance abuse and tobacco dependency<br><br>c) medical or behavioural treatments targeting populations with a specific condition, including cognitive behavioural therapy or exercise.<br><br>*Note: 3ie has traditionally not included studies of types a-c in our repositories. But we are reviewing this policy, as it is out of sync with current thinking in development. For the time being we are not including these studies, but we are exploring whether we have the resources to conduct searching and screening to identify studies of this type that belong in our repositories.* | | | |
| | **IF YES, THEN EXCLUDE** | | | |

## Appendix A: Minimum reporting requirements for studies to be included in the IER

| **Randomised evaluations** |
|---|
| |
| **Definition** |
| An impact evaluation design in which random assignment has been used to allocate the intervention amongst members of the eligible population. |
| |
| **Reporting requirements for establishing a counterfactual** |
| Clear description of random *assignment* process |
| |
| **Requirements for reporting results** |
| Post-intervention differences between groups or sub-populations should be calculated using method of data analysis such as single difference or ordinary least squares (OLS). |
| Results can be reported as odds ratios or confidence intervals. |
| Statistical tests for significance are required. |
| |
| **Regression discontinuity design** |
| |
| **Definition** |
| An impact evaluation design in which the treatment and comparison groups are identified as being those just on either side of a threshold value of a variable. |
| |
| **Reporting requirements for establishing a counterfactual** |
| The threshold is clearly defined. |
| Established continuity at threshold |
| Distribution of covariates and outcome measures around threshold is compared to ensure 'balance'. |
| |
| **Requirements for reporting results** |
| Post-intervention differences between groups should be calculated using method of data analysis such as single difference or OLS. |
| Results can be reported as odds ratios or confidence intervals. |
| Statistical tests for significance are required. |
| |
| **Statistical matching (PSM and others)** |
| |
| **Definition** |
| An impact evaluation design in which the comparison group is constructed using statistical matching techniques, such as propensity scores. A propensity score is the probability of participating in the intervention, as given by a probit regression on observed characteristics. |
| |

| |
|---|
| **Reporting requirements for establishing a counterfactual** |
| Covariates used to estimate propensity score are clearly listed. |
| The authors test the quality of the matching procedures using one of the following tests: |
| a) Covariate balanced comparison test before and after the matching |
| b) Histogram of propensity score before and after the matching |
| c) Pseudo $R^2$ before and after the matching |
| d) Sensitivity analysis to address the issue of hidden bias related to unobservable variables |
| |
| **Requirements for reporting results** |
| Post-intervention differences between groups should be calculated using method of data analysis such as single difference or OLS. |
| Results can be reported as odds ratios or confidence intervals. |
| Statistical tests for significance are required. |
| |
| **Difference-in-differences and fixed effects estimation** |
| |
| **Definition** |
| Difference in differences calculates the change in the outcome observed in the treatment group compared to the change observed in the comparison group. Fixed effects, when using panel data, control for time-invariant characteristics by exploring the relationship between the dependent and explanatory variables *within* an entity (e.g. individual, household and so on). |
| |
| **Requirement for testing assumptions** |
| To test parallel trends assumptions, the paper must meet at least one (1) of the criteria below: |
| [1] Use at least two serial observations on the treatment and comparison groups before the start of the programme. This means that the evaluation would require three serial observations: two pre-intervention observations to assess the preprogramme trends, and at least one post-intervention observation to assess impact with the difference-in-difference method. |
| [2] Perform a 'placebo test' by conducting an additional difference-in-difference estimation using a 'fake' treatment group: that is, a group that you know was not affected by the programme. |
| [3] Perform the placebo test not only with a 'fake' treatment group, but also with a 'fake' outcome. |
| [4] Perform the difference-in-difference estimation using different comparison groups. |
| See Gertler et al.'s handbook (pages 137-138) for a deeper explanation of the tests below |
| |
| **Requirements for reporting results** |
| Post-intervention differences between groups should be calculated using a regression with a time X treatment interaction. |
| Statistical tests for significance are required. |
| |

| **Instrumental variable estimation** |
| --- |
| |
| **Definition** |
| The IV method is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomised experiment. A valid instrument induces changes in the explanatory variable but has no independent effect on the dependent variable. The explanatory variable only affects the dependent variable through the instrument. |
| |
| **Requirement for testing assumptions** |
| Test underlying assumptions: |
| Theoretical discussion on why the instrument is correlated with the explanatory variable and not with the outcome variable or error term. |
| Instrument must meet the relevance condition: authors should test for significant correlation between instrument and explanatory variable |
| |
| **Requirements for reporting results** |
| Statistical tests for significance are required. |

## Appendix B: Countries by income status

| LOW- AND MIDDLE-INCOME COUNTRIES (L&MICs) | | | |
|---|---|---|---|
| Afghanistan | Eritrea | Marshall Islands | Sudan |
| Albania | Ethiopia | Mauritania | Suriname |
| Algeria | Fiji | Mauritius | Swaziland |
| Angola | Gabon | Mexico | Syrian Arab Republic |
| Armenia | Gambia, The | Micronesia, Federal States | Tajikistan |
| Azerbaijan | Georgia | Moldova | Tanzania |
| Bangladesh | Ghana | Mongolia | Thailand |
| Belarus | Grenada | Montenegro | Timor-Leste |
| Belize | Guatemala | Morocco | Togo |
| Benin | Guinea | Mozambique | Tonga |
| Bhutan | Guinea-Bissau | Myanmar | Tunisia |
| Bolivia | Guyana | Namibia | Turkey |
| Bosnia and Herzegovina | Haiti | Nepal | Turkmenistan |
| Botswana | Honduras | Nicaragua | Tuvalu |
| Brazil | India | Niger | Uganda |
| Bulgaria | Indonesia | Nigeria | Ukraine |
| Burkina Faso | Iran, Islamic Republic | Pakistan | Uzbekistan |
| Burundi | Iraq | Palau | Vanuatu |
| Cambodia | Jamaica | Panama | Vietnam |
| Cameroon | Jordan | Papua New Guinea | West Bank and Gaza |
| Cape (Cabo) Verde | Kazakhstan | Paraguay | Yemen, Republic |
| Central African Republic | Kenya | Peru | Zambia |
| Chad | Kiribati | Philippines | Zimbabwe |
| China | Korea, Democratic Republic | Romania | |
| Colombia | Kosovo | Rwanda | |
| Comoros | Kyrgyz, Republic | Samoa | |
| Congo, Democratic Republic | Lao PDR | São Tomé and Principe | |
| Congo, Republic | Lebanon | Senegal | |
| Costa Rica | Lesotho | Serbia | |
| Côte d'Ivoire (Ivory Coast) | Liberia | Sierra Leone | |
| Cuba | Libya | Solomon Islands | |
| Djibouti | Macedonia, FYR | Somalia | |
| Dominica | Madagascar | South Africa | |
| Dominican Republic | Malawi | South Sudan | |
| Ecuador | Malaysia | Sri Lanka | |
| Egypt, Arab Republic | Maldives | St. Lucia | |
| El Salvador | Mali | St. Vincent and the Grenadines | |

| FORMER LOW- AND MIDDLE-INCOME COUNTRIES |
|---|
| Czechoslovakia |
| Gibraltar (Developed: 2009-2010) |
| Mayotte (Developed: 1990) |
| Netherlands Antilles (Developed: 1994-2009) |
| Serbia and Montenegro |
| USSR |
| Yugoslavia |

| TRANSITIONAL COUNTRIES | | |
|---|---|---|
| **Name** | **L&MIC period** | **High-income country period** |
| American Samoa | 1990-present | 1987-1989 |
| Antigua and Barbuda | 1987-2001; 2003-2004; 2009-2011 | 2002; 2005-2008; 2012-present |
| Argentina | 1987-2013; 2015-present | 2014 |
| Aruba | 1991-1993 | 1987-1990; 1994-present |
| Bahrain | 1990-2000 | 1987-1989; 2001-present |
| Barbados | 1987-1988; 1990-1999; 2001; 2003-2005 | 1989; 2000; 2002; 2006-present |
| Chile | 1987-2011 | 2012-present |
| Croatia | 1992-2007; 2016-present | 2008-2015 |
| Cyprus | 1987 | 1988-present |
| Czech Republic | 1992-2005 | 2006-present |
| Equatorial Guinea | 1987-2006 | 2007-present |
| Estonia | 1991-2005 | 2006-present |
| Guam | 1990-1994 | 1987-1989; 1995-present |
| Greece | 1987-1995 | 1996-present |
| Hungary | 1987-2006; 2012-2013 | 2007-2011; 2014-present |
| Isle of Man | 1990-2001 | 1987-1989; 2002-present |
| Latvia | 1991-2008; 2010-2011 | 2009; 2012-present |
| Lithuania | 1991-2011 | 2012-present |
| Macao (SAR) | 1987-1993 | 1994-present |
| Malta | 1987-1988; 1990-1997; 1999; 2001 | 1989; 1998; 2000; 2002-present |
| New Caledonia | 1987-1994 | 1995-present |
| Northern Mariana Islands | 1992-1994; 2002-2006 | 1995-2001; 2007-present |
| Oman | 1987-2006 | 2007-present |
| Poland | 1987-2008 | 2009-present |
| Portugal | 1987-1993 | 1994-present |
| Puerto Rico | 1987-1988; 1990-2001; | 1989; 2002-present |
| Republic of Korea | 1987-1994; 1998-2000 | 1995-1997; 2001-present |
| Russia | 1991-2011; 2015-present | 2012-2014 |
| Seychelles | 1987-2013 | 2014-present |
| Slovak Republic | 1992-2006 | 2007-present |
| Slovenia | 1992-1996 | 1997-present |
| Saudi Arabia | 1990-2003 | 1987-1989; 2004-present |

| TRANSITIONAL COUNTRIES | | |
|---|---|---|
| **Name** | **L&MIC period** | **High-income country period** |
| St. Kitts and Nevis | 1987-2010 | 2011-present |
| Trinidad and Tobago | 1987-2005 | 2006-present |
| Uruguay | 1987-2011 | 2012-present |
| Venezuela | 1987-2013; 2015-present | 2014 |

| HIGH-INCOME COUNTRIES | |
|---|---|
| Andorra | Australia |
| Austria | Bahamas |
| Belgium | Bermuda |
| Brunei Darussalam | Canada |
| Cayman Islands | Channel Islands |
| Curacao | Denmark |
| Faeroe Islands | Finland |
| France | French Polynesia |
| Germany | Greenland |
| Hong Kong (SAR) | Iceland |
| Ireland | Israel |
| Italy | Japan |
| Kuwait | Liechtenstein |
| Luxembourg | Monaco |
| Netherlands | New Zealand |
| Norway | Qatar |
| San Marino | Singapore |
| Sint Maarten (Dutch Part) | Spain |
| St. Martin (French Part) | Sweden |
| Switzerland | Taiwan |
| Turks and Caicos Islands | United Arab Emirates |
| United Kingdom | United States |
| Virgin Islands (US) | |

## Appendix C: Efficacy

Consider including efficacy studies if any one of the following criteria are met:

1. The intervention being evaluated promotes a social, economic or behavioural change either as one of the final measured outcomes or as a mechanism within the theory of change (beyond the self-administration of a drug). For example, the study may include health behaviour messaging, training, provision of information, or screening or surveillance for specific disease conditions.
2. The study measures any other outcomes in addition to or beyond purely biomedical indicators (e.g. returns to education, economic productivity, quality of life, disability adjusted life years [DALYs] and spillover effects).
3. The study records any additional formative information that could guide the design or execution of future studies. For example, an RCT that also measures acceptability of a particular treatment (measuring respondent satisfaction with treatment not merely a rate of compliance or uptake) would be included.
4. The treatment is both prepared and delivered by a community health worker, or trained layperson (e.g. parent, teacher or community member and not merely one of the programme or study enumeration team).
5. The programme or outcomes measured answer, or attempt to answer, a question relevant to the roll-out of international development policies or interventions.

For more information: https://www.3ieimpact.org/blogs/efficacy-effectiveness-continuum-and-impact-evaluation