International Initiative for Impact Evaluation

Push button replication

Project Description

June 2016

Background

The International Initiative for Impact Evaluation (3ie) seeks to improve the lives of poor people in low- and middle-income countries by providing and summarizing evidence about which development interventions work, when, why, and for how much. In addition to providing and summarizing such evidence, 3ie seeks to improve the quality and reliability of evidence that is produced and used for development policy making. To this end, 3ie established the Replication Programme for impact evaluations of development policies and programmes.

In May 2015, 3ie gathered a group of critics, supporters, and people generally interested in replication research for a one day consultation event in Washington, DC (summary here). Somewhat to our surprise, one of the livelier discussions at the event centered on whether it is reasonable to expect that the vast majority of published empirical studies can be exactly reproduced. In particular, the expectation was that original data and programming code or coding documentation from an existing article could be used by a third party to easily reproduce the published results.

All present agreed that this kind of reproduction, where one can essentially push a button and recreate the published tables in an article, is the most basic replication question. Some argued that this expectation should be a given – that of course original authors always have the data and programming documentation to reproduce their work. Others expressed strong doubts about how frequently authors really can provide the required materials to reproduce the published findings. These doubters argued that replication research should focus, either primarily or additionally, on this very first line of verification, at least until the relevant fields reach a point where we can take exact reproduction of published results for granted.

Objectives

The primary objective of the PBR project is to establish procedures and standards for push button replication so that original authors and replication researchers can better align their expectations and actions around this third-party verification process. The second objective is to test whether development impact evaluations, those studies using experimental and quasi-experimental

Washington, DC
1029 Vermont Avenue, NW, Suite 1000, Washington, DC 20005, United States of America
3ieus@3ieimpact.org Tel: +1 202 629 3939

methods to evaluate the effectiveness of development programmes in low- and middle-income countries, are generally push button verifiable.

The push button replication concept

PBR research attempts to confirm the validity of published results using both the original data and the programming code from a study. The premise behind a PBR study is that the third party researcher should not need to make any significant adjustments, write new code, or conduct additional analysis in order to arrive at the published results. A PBR is thus a step before pure replication. While a pure replication asks the question can we reproduce the published results using the original data and the methods described in the original study, a PBR asks the question can we use the original authors' programming code with the original data to reproduce the published results. Pure replication studies can uncover errors where the programming code incorrectly implements the estimation methods described. PBR studies reveal cases where there are no or insufficient data and code or where the programmes do not run.

Programme features

This document outlines the features of the PBR project.

Candidate studies list

Candidate studies for PBR must be impact evaluations of social and economic development programmes in low- and/or middle-income countries. According to the 3ie Principles for Impact Evaluation, high-quality impact evaluations measure the net change in outcomes amongst a particular group, or groups, of people that can be attributed to a specific programme using the best methodology available, feasible and appropriate to the evaluation question(s) being investigated and to the specific context. Social and economic development programmes have, as their primary purpose, the improvement of a population's quality of life, whether they are financed by public agencies, NGOs, or private institutions; for example programmes aimed at improving health, education, employment, access to credit, infrastructure, and reducing income-poverty.

The sample of articles is selected from all the development impact evaluations published in 2014 in the top ten journals for these studies. We determined the top ten journals as those that published the greatest number of development impact evaluations during the period 2010 through 2012 as catalogued in 3ie's comprehensive impact evaluation repository. These journals span several disciplines including public health, political science, economics, and others. These top ten journals are:

- AIDS and Behavior
- American Economic Journal
- The BMJ
- Economic Development and Cultural Change
- Journal of Development Economics
- Journal of Development Effectiveness
- Plos One
- The Lancet

- Tropical Medicine and International Health
- World Development

Taking all development impact evaluations published by these ten journals in 2014 yields a sample of 123 studies (excluding pre-print online versions). The 2014 sample allows us to examine whether the recent emphasis on research transparency across several disciplines has translated into improvements in push button replicability – both in terms of the willingness and ability of original authors to provide data and code and in the success of replications where the files are provided. The narrow window between the publication date for our sample and when our project is conducted should provide an upper bound for the expected push button replicability of this type of research, as it should be less burdensome and more expected for authors to provide data, code, and replication instructions for more recent research.

PBR communication and data acquisition

The PBR protocol begins with instructions for how to reach out to original authors. If the data and code are public, PBR researchers will email the original authors to inform them that they will be conducting a PBR, as well as inquiring on the statistical software and specific version used by the authors.

If the data and code are not publicly available, the PBR researchers will request them from the original authors. This email will include a brief explanation of the PBR project and the request of any materials needed to replicate tables and figures from the original publication including: code, data, replication instruction, and the statistical software and specific version used in the analysis. PBR researchers will continue to follow up with original authors after two weeks. If the original authors ask for more than three months to provide the data or do not respond, the PBR will be classified as no access.

PBR process

PBR researchers will replicate the results using the existing code and data sets. When available, PBR researchers should follow any replication instructions provided by the original authors. If PBR researchers are unable to run code sent by the original researchers, they will confirm the data and code with the original authors. PBR researchers will troubleshoot minor complications if they observe them.

Once they have run the PBR, researchers will compare the replication results and the original paper and report the results. The PBR protocol includes a checklist and report template for documenting the process and results of a push button replication in order to encourage consistency between the PBR research reports. It also provides a typology for classifying the results of a push button replication. The PBR researcher includes the justification for the selected classification in the report.

All of the PBR project materials and the reports from all 3ie-funded PBR studies will be made available on the Open Science Framework platform. The protocol includes instructions for when and how results are made available to the original authors before they become public.