

International Initiative for Impact Evaluation



WORKING PAPER 6

**A checklist for the reporting of randomized
control trials of social and economic policy
interventions in developing countries: CEDE
Version 1.0**

Ron Bose
January 2010

About 3ie

The International Initiative for Impact Evaluation (3ie) works to improve the lives of people in the developing world by supporting the production and use of evidence on what works, when, why and for how much. 3ie is a new initiative that responds to demands for better evidence, and will enhance development effectiveness by promoting better informed policies. 3ie finances high-quality impact evaluations and campaign to inform better program and policy design in developing countries.

3ie Working Paper series covers both conceptual issues related to impact evaluation and findings from specific studies or synthetic reviews.

This Working Paper was edited by Dr. Howard White, 3ie Executive Director.

© 3ie, 2009

Contacts

International Initiative for Impact Evaluation
c/o Global Development Network
Post Box No. 7510
Vasant Kunj P.O.
New Delhi – 110070, India
Tel: +91-11-2613-9494/6885
www.3ieimpact.org

A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries: CEDE Version 1.0

*Ron Bose, International Initiative for Impact Evaluation **

Abstract

The consolidated standards of reporting trials (CONSORT) checklist were developed to assist investigators, authors, reviewers, and journal editors provide the necessary information to be included in reports of controlled medical trials. We augment the CONSORT reporting by adapting and elaborating the checklist to the context of trials of development interventions. We call this revised list the CONSORT Elaborations for Development Effectiveness or CEDE. This checklist emphasizes the reporting of underlying theories and descriptions of intervention and comparison conditions, research design, and detailed discussion of the protocol to mitigate the threats to the randomized evaluation design of studies. Systematizing, and greater transparency, in the reporting formats for RCTs will enable the community of evaluators, policy makers, program officer to be privy to the many steps in an RCT implementation, and to better judge the internal and external validity of specific RCTs, both absolutely and relative other methods of evaluation. The CEDE checklist is not meant to be the basis for evaluation of the RCT methodology, but to promote better reporting of data reporting of published and completed studies. These guidelines should evolve alongside the state of the art of the field of experimental trial designs for the evaluation of social and economic policy interventions.

Keywords: CONSORT statement; development effectiveness; impact evaluation; randomized controlled trials

* Special thanks are due to David McKenzie for sharing a copy of his joint paper with Miriam Bruhn, Chris Blattman, Marc Shotland (J-PAL) and Chris Woodruff for comments on an earlier draft of this paper, and to Howard White for instigating the project. All errors however remain my own. The views expressed in this paper may not be taken to represent those of 3ie nor any of its member organizations. Contact email: cede@3ieimpact.org.

Creating a culture in which rigorous randomised evaluations are promoted, encouraged, and financed has the potential to revolutionise social policy during the 21st century, just as randomised trials revolutionised medicine during the 20th

Esther Duflo, co-Founder of Poverty Action Lab

I. Introduction

Growing calls for evidence-based policy making, led in part by the limited state of knowledge of what works and does not work in the field of development, have served as catalyst for increased demand for rigorous impact evaluation of social programs. High quality rigorous impact evaluations have been defined as those that enable the evaluator to accurately ascribe changes in outcomes of interest to specific policy interventions or initiatives.ⁱ Attribution of program impacts to specific inputs most usually requires the analyst to construct the counterfactual, what would the individual (or village, or region) have looked like in the absence of the policy intervention? Evaluators adopting this approach begin by constructing a relevant control group which is observed and compared against the group of program beneficiaries (the treatment group).ⁱⁱ

One approach to creating a comparable group for the purposes of evaluation is randomization. The rationale behind the approach is that random allocation of the treatment will, with a sufficiently large sample size, ensure that treatment and control (untreated) groups are comparable in all respects except the one being studied, i.e. the intervention being evaluated. The approach is not to be confused with *ex post* random samples of the treated and untreated populations. The latter will, in the absence of some credible matching strategy, be subject to selection bias. This bias is in principle avoided when the treatment is randomly allocated *ex ante*. Proper reporting of the trial – the point of these guidelines – allows judgment of the extent to which bias has indeed been successfully avoided.

The advantage of randomization over other matching strategies arises from the possible presence of time varying, unobservable determinants of program participation which are correlated with the outcomes of interest. Whilst evaluators can achieve comparability on factors that are known to influence the outcome, such as age, sex, race, or severity of disease, by matching for these factors, they cannot match persons for factors whose influence is not known, or cannot be measured. Hence the problem of unobserved characteristics is addressed by the random assignment of individuals (or households, firms, schools or whatever is the unit at which random assignment takes place) to the treatment and comparison groups. Randomisation helps to ensure that the distribution of all factors - known and unknown, measurable and not measurable - is based on chance and not some factor, such as participant preference, that may lead to a bias in assignment. In addition, randomization is the means by which the evaluator avoids introducing conscious and subconscious bias into the process of allocating individuals to the treatment or comparison groups. Preventing biased assignment permits causal inference in the interpretation of program impacts. Accordingly well-designed and implemented randomized controlled trials are considered by some to be the "gold standard" for program evaluation. Randomised trials for the evaluation of economic and social policies have certainly not been as common as they are in fields of medicine, and in the past were largely been confined to developed nations, with the US being in the vanguard, to a lesser extent in the UK and Australia. Now with the work of researchers associated with the Poverty Action Lab (quoted at the start of this paper) and Innovations in Poverty Action (IPA), the portfolio of studies financed under the World Bank's Development Impact Evaluation Initiative

(DIME), and Spanish Impact Evaluation Fund (SIEF), and the financing being made available by the International Initiative for Impact Evaluation (3ie) mean there has been growing interest within the development community in these methods for evaluation in the developing context (White, 2009a).

In clinical medicine, the Consolidated Standards of Reporting Trials (CONSORT) statement were developed to assist investigators, authors, reviewers, and editors as to the necessary information to be included in reports of controlled clinical trials. Better documentation and complete reporting of trials encouraged investigators to modify their clinical practice in response to the best available evidence (Moher et al 2001). The CONSORT statement was first published in 1996 and revised in 2001. This statement consists of a checklist and flow diagram to guide writers and reviewers on the information that should be available from published reports of randomized control trials (ibid). The statements specifies a set 22 items required for a clear and transparent account of what was done in a research study, and what was found, reflecting in particular issues that might introduce bias into the research. Today the CONSORT statement, based on consensus opinion of experts in the field of health, including research methodologists and journal editors, has been credited with improved the quality of clinical RCTs , and increasingly endorsed by many leading medical journals, editorial associations, professional societies, and funding agencies (ibid).

To the best of our knowledge no such framework exists in the design of experimental evaluations of economic and social policy interventions for improving development effectiveness. To this end this paper is intended to promote discussion on raising the quality of RCTs in the development arena by borrowing and building upon the experience from initiatives that aimed at raising the quality of clinical trials through better and complete reporting. Harmonization of reporting of RCT evaluations of social and economic policies in the developing context will reduce the variability in the quality and reliability of finding from RCTs. This improvement will in turn ensure that the community of development evaluators, policy makers, donors, NGOs and broader civil society have a better understanding of, and insight to, the many steps in RCT design/implementation. They will thus be better able to judge the internal and external validity of RCTs relative other methods of policy evaluation.

II. CONSORT Extensions to evaluations of social and economic policy interventions

We provide elaborations on CONSORT checklist items for reporting of RCTs in social and economic policies in developing countries. In particular the proposed checklist extensions emphasize the statement of the program theory underlying the behavioral change stemming from the proposed intervention settings and context; full reporting of outcomes; and inclusion of information related to the implementation of randomized design needed to assess possible biases in the outcome data. These extensions are presented in Table 1. In all other cases readers are referred to the elaboration on the items provided on from the CONSORT website (<http://www.consort-statement.org>). The checklist is accompanied by Figure 1, which documents participant flow through a trial.

III. DISCUSSION

In compiling this checklist and the related extensions we recognize several challenges in promulgation, acceptance and use of these reporting standards for randomized intervention evaluations. One possible concern would be space limitations. Many journals have required page limits toward shorter rather than longer articles. In light of these, we would recommend using the journal website where data and codes already stored to be the venue. Alternatively, full study details may be recorded on a project website (referred to in the paper), or the website of a relevant international organization, including 3ie .

It should be reiterated that the CEDE checklist is only a suggested set of guidelines which should be considered a work in progress. It is likely that improvements will be necessary; moreover, adaptations may be needed to refine the standards for specific fields of intervention research, and additional specifications for specific types of randomized evaluations. Comments can be sent to: CEDE@3ieimpact.org. Journal editors, heads of evaluation in major international and national agencies will be encouraged to endorse this effort by publishing editorials or commentaries on the CEDE statement or by referencing it in their publication guidelines for authors and reviewers. To increase accessibility and ease of use, the updated versions of the CEDE statement will be posted on the 3ie website (<http://www.3ieimpact.org>)

IV. CLOSING REMARKS

A recent paper (Blattman 2008), characterizes the current spate of development evaluation employing a randomized evaluation design as Evaluation 1.0, i.e. these were studies focused primarily on quantifying the (average) impacts of program interventions. These are distinguished from next generation of enhanced evaluations, Evaluation 2.0, which are RCTs whose focus is as much on “how” as it is on “what” works, hence on unpacking the causal chain from inputs to processes to intermediate and final outcomes to reveal how programs work (do not work). This is done by using evaluation design as a tool to elucidate the manner in which people, processes and program elements in the intervention combine to produce the observed impacts or changes in the outcome(s) of interest (see also White, 2009b). Such types of evaluation improve not just the capacity of undertaking experimental assessment, and but yield better guidance on how program scale ups, changes in the program/evaluation design would be expected to affect impacts over time and in other cultural/political and social milieus. It is hoped that standardized and transparent reporting of trials, through the aid of the CEDE statement, will be able to do just that, i.e. allow policy makers, evaluators, program staff and program managers to accurately appraise the validity and judge implications of the findings studies relative to non-randomized evaluation methods.

Figure 1 Participant Flow Template

Source- <http://www.consort-statement.org/>

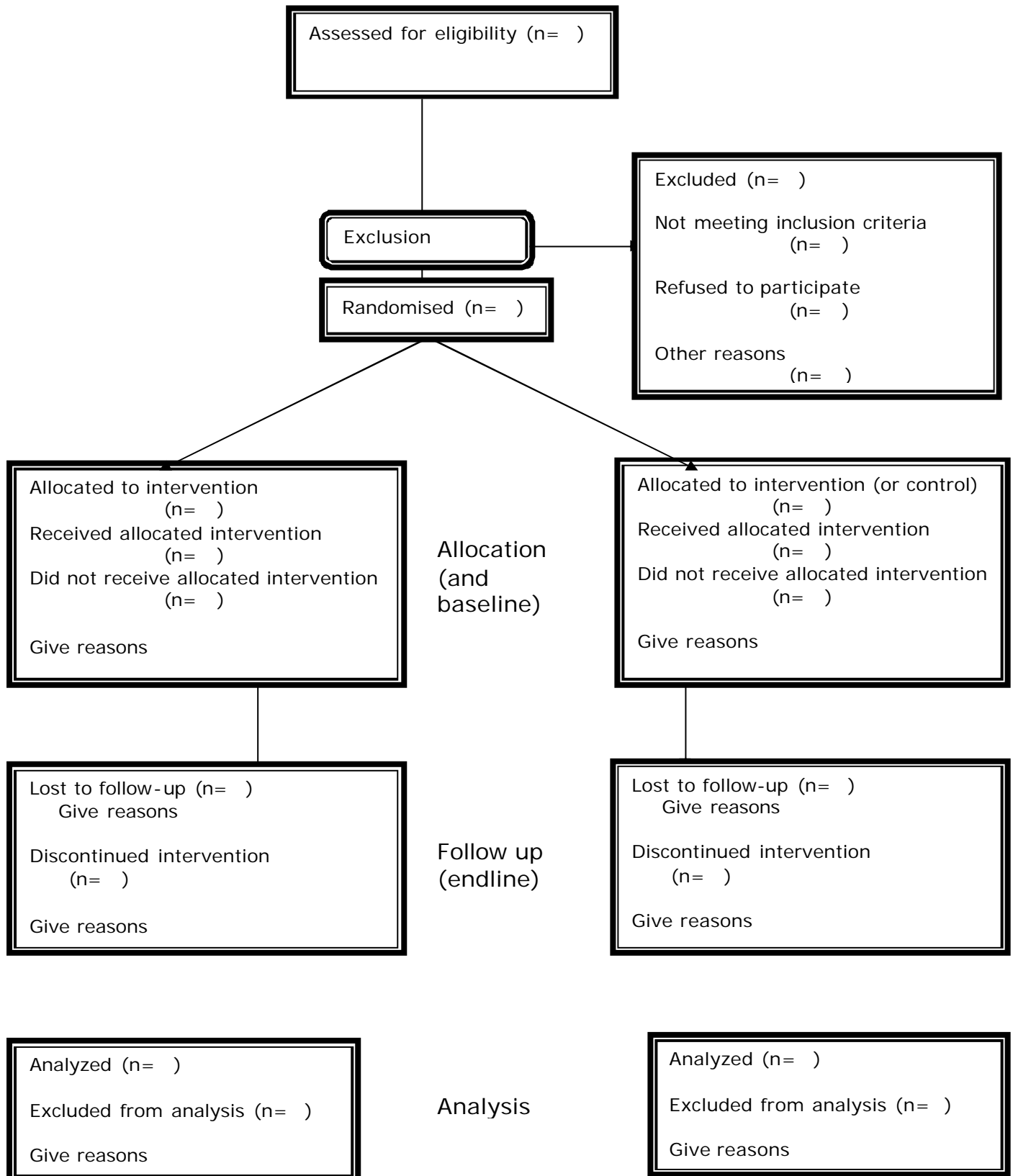


Table 1 - CONSORT Extensions Checklist (CEDE)

Original CONSORT	Item	Description from CONSORT	Extension	Examples
Title and abstract	1	How participants were allocated to the intervention	<p>The title of the paper should indicate that the results being reported are from a randomized control trial.</p> <p>The abstract should include information on how units were allocated to interventions along with a structured abstract recommended</p>	<p>'Small individual loans and mental health: a randomized controlled trial among South African adults' (Fernald et al., 2008)</p> <p>'We explore this argument through a field experiment in Kenya, in which we randomized the price at which prenatal clinics could sell long lasting anti-malarial insecticide-treated nets (ITNs) to pregnant women' (Cohen and Dupas, 2007)</p>
Background	2	Scientific background and rationale for undertaking the study	Clear statement of the new or traditional theories used in designing behavioral interventions and clear justification for undertaking the study based on systematic reviews of literature to identify evaluation gaps	'researchers ... found considerable heterogeneity, with many micro-entrepreneurs (in particular females) earning negative returns to capital... This calls into the question the "poor but rational" view that micro-entrepreneurs maximize profits subject to their financial constraints' (Karlan and Valdivia, 2009)
<i>Methods</i>				
Participants	3	Eligibility criteria for participants; settings and locations where the data were collected	<p>a) Eligibility criteria for participants and clusters and the settings and locations including criteria at different levels in recruitment /sampling plan (e.g., cities, villages, subjects) with a clear description of how the established theories and concepts were maintained as it relates to the participant inclusion criteria</p> <p>b) Methods of recruitment (referrals, self selection) including the sampling method if a systematic sampling plan was implemented</p>	'Using the 2001 Sri Lankan Census, we selected 25 Grama Niladhari divisions (GNs) in three Southern and South-Western districts of Sri Lanka: Kalutara, Galle and Matara. A GN is an administrative unit containing on average around 400 households. We used the Census to select GNs with a high percentage of own-account workers and modest education levels, since these were most likely to yield enterprises with invested capital below the threshold we had set. GNs were also stratified according to the degree of exposure of firms to the December 26, 2004 Indian Ocean tsunami' (De Mel et al 2009)

Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered	<p>Details of the interventions intended for each study condition and how and when they were actually administered, specifically including:</p> <ul style="list-style-type: none"> a) Content: what was given? b) Delivery method: how was the content given? c) Unit of delivery: how were subjects grouped during delivery? d) Deliverer: who delivered the intervention? e) Setting: where was the intervention delivered? f) Exposure quantity and duration: how many sessions or episodes g) Time span: how long was it intended to take to deliver the intervention h) Activities to increase compliance or adherence (e.g., incentives) i) The underlying logic by which the intervention is expected to affect the intended outcome j) Whether theory or experience elsewhere suggest unanticipated outcomes to be included in the trial 	<p>'Our experiment targeted ITN distribution to pregnant women visiting health clinics for prenatal care.¹⁰ We worked with 20 rural public health centers chosen from a total of 70 health centers in the region, 17 of whom were private and 53 were public... We randomly assigned them to one of five groups: 4 clinics formed the "control group"; 5 clinics were provided with ITNs and instructed to give them free of charge to all expectant mothers coming for prenatal care; 5 clinics were provided with ITNs to be sold at 10 Ksh (corresponding to a 97.5 percent subsidy); 3 clinics were provided with ITNs to be sold at 20 Ksh (95.0 percent subsidy); and the last 3 clinics were provided with ITNs to be sold at 40 Ksh (90 percent subsidy). Clinics were provided with financial incentives to carry out the program as designed. For each month of implementation, clinics received a cash bonus (or a piece of equipment of their choice) worth 5,000 Ksh (approximately \$75) if no evidence of "leakage" or mismanagement of the ITNs or funds was observed. Clinics were informed that random spot checks of their record books would be conducted, as well as visits to a random sub-sample of beneficiaries to confirm the price at which the ITNs had been sold and to confirm that they had indeed purchased an ITN (if the clinic's records indicated so). Despite this, we observed leakages and mismanagement of the ITNs in 4 of the 11 clinics that were asked to sell ITNs for a positive price. We did not observe any evidence of mis-management in the five clinics instructed to give out the ITNs for free... The ITN distribution program was phased into program clinics between March and May 2007, and was kept in place for at least 3 months in each clinic,</p>
---------------	---	---------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

				<p>throughout the peak “long rains” malaria season and subsequent months. Posters were put up in clinics to inform prenatal clients of the price at which the ITNs were sold. Other than offering a free hemoglobin test to each woman on survey days, we did not interfere with the normal procedures these clinics use at prenatal care visits, which in principle include a discussion of the importance of bed net usage.’ (Cohen and Dupas, 2007: 9-10)</p>
Objectives	5	Specific objectives and hypotheses	Clear statement of the objectives of the intervention, intended outcomes (see item 6), and related hypotheses.	<p>‘To the extent that students benefit from having higher-achieving peers, tracking [streaming] students into separate classes by prior achievement could disadvantage low-achieving students while benefiting high-achieving students, thereby exacerbating inequality... On the other hand, tracking could potentially allow teachers to more closely match instruction to students’ needs, benefiting all students. This suggests that the impact of tracking may depend on teachers’ incentives.’ (Duflo et al., 2009: 2)</p>
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors)	<p>Precise definition of outcome indicators; related survey questions may often usefully be included in an annex. Expected differential impact by sub-group should have been specified prior to undertaking the analysis (see item 17).</p>	<p>‘The second part of the survey consisted of a series of tests: a standardized math test, a cognitive test, and a multiple choice psychological test. The math test was based on part of a test that was originally created for The Trends in International Mathematics and Science Study (TIMSS). The cognitive test was developed to test for cognitive reasoning ability and was developed by the World Bank. The psychological test of well-being, the Mental Health Test (MHT), was developed by Professor Bucheng Zhou. The Zhou MHT contains 100 yes/no questions. Lower test scores, correspond to a healthier mental state.’ (REAP, 2009)</p>

Sample size	7	How sample size was determined; explanation of any interim analyses and stopping rules when applicable	<p>a) The power calculation its underlying assumptions for the trial would need to be specified. If a formal power calculation was used the authors should identify the primary outcome on which the power calculation was based (see Item #6), making specific reference to the previous literature findings, <u>and</u>, allowing (or describing) for lower levels of effectiveness based on the size of the intervention.</p> <p>b) Software used to compute Power Calculations</p>	Power calculations were done using STATA v. 10 for multi-site cluster randomized control trial. Effects size, intra-cluster correlation, number of sites and cluster size assumed were based on review of previous literature' (Duflo et al, 2008)
Randomization: sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification)	<p>a) Unit of assignment (the unit to which the treatment is being assigned), e.g., individual, firm, school.</p> <p>b) Method used to assign units to study conditions, including details of any restriction (e.g., blocking, stratification, and minimization). For stratification the number of strata used should be clearly described</p>	'In our samples of 30 observations, we stratify on 2 variables, forming 8 strata. In the samples of 100 and 300 observations, we also stratify on 3 variables (24 strata), and also on 4 variables (48 strata)' (Bruhn and Mckenzie, 2008)
Randomization: allocation concealment	9	Method of implementing the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was	How was randomization done (coin toss, random number generator)? What software was used to generate the random number generator?	'The randomization was a simple piece of Windows software that included a data entry screen, where officers inputted client information, and then were presented with a randomization results screen. Random assignment to the Treatment condition constituted being part of a group of applications for which the Lender received "encouragement to reconsider" (i.e. to take a "second look"); those

concealed until interventions were assigned

with better credit scores among the marginal rejects were treated with probability 0.50, and those with worse credit scores among the marginal rejects were treated with probability 0.25. The treated group did not receive "randomized approval" for the loan because loan officers had pecuniary incentives to be risk-averse, and the Lender deemed it impractical to force officers to comply strictly with the randomizer's decision' (Fernald et al 2008)

Randomization: implementation

10

Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups

The study should

- a) Clearly explain who performed the randomization
- a) Was the randomization carried out in public or private?

Statistical methods

11

Statistical methods used to compare groups for primary outcomes; methods for additional analyses, such as subgroup analyses and adjusted analyses

- a) Methods for imputing missing data, if any
- b) Statistical software or programs used
- c) If the unit of analysis differs from the unit of assignment, the analytical models were used to account for this (e.g., adjusting the standard error estimates by the design effect or using multilevel analysis)

As analyses were performed at the individual level and randomization was done at the community level, a prior estimate of the intraclass correlation coefficient was used to adjust the standard error estimates before calculating confidence intervals

Results

Participant flow	12	<p>Flow of participants through each stage (a diagram is strongly recommended – see Figure 1)—specifically, for each group, report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcomes</p>	<p>a) Flow of participants through each stage of the study: enrollment, assignment, allocation and intervention exposure, follow-up, analysis (a diagram is strongly recommended)</p> <ul style="list-style-type: none"> i. Enrollment: the numbers of participants screened for eligibility, found to be eligible or not eligible, declined to be enrolled, and enrolled in the study ii. Assignment: the numbers of participants assigned to a study condition iii. Allocation and intervention exposure: the number of participants assigned to each study condition and the number of participants who received each intervention iv. Follow-up: the number of participants who completed the follow-up or did not complete the follow-up (i.e., lost to follow-up), by study condition v. Analysis: the number of participants included in or excluded from the main analysis, by study condition <p>b) Description of deviations from study protocol, along with reasons</p> <p>The potential for differential adherence and follow up is exacerbated in the cluster</p>
------------------	----	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

randomized design because there are two levels at which drop-outs can occur: whole clusters or individuals in a cluster. It is therefore important to describe the flow of both clusters and individuals when reporting a cluster randomized trial. A flow diagram is usually the best way to present this information (Figure 1).

Recruitment	13	Dates defining the periods of recruitment and follow-up	Provide all relevant dates (year and month) of both the intervention and the data collection	
Baseline data	14	Baseline demographic and clinical characteristics of each group	<p>a) Baseline characteristics for each study condition relevant</p> <p>b) Baseline comparisons of those lost to follow-up and those retained, overall and by study condition</p> <p>c) Comparison between study population at baseline and target population of interest</p>	The treatment and control groups did not statistically differ with respect to demographic data (gender, age, race/ethnicity; $P > .05$ for each), but the intervention group reported a significantly greater baseline incidence of child diarrhea ($P = .03$)
Numbers analyzed	15	Number of participants (denominator) in each group included in each analysis and whether analysis was by 'intention-to-treat' or 'treatment of the treated'; state the	If analysis was <u>not</u> intent to treat (ITT) explain clearly how non-compliers were treated in the statistical analysis	The primary analysis was <u>intention to treat</u> and included all subjects as assigned with available 9-month outcome data (125 of 176 assigned to the intervention and 110 of 164 assigned to the control)

results in absolute numbers when feasible (e.g. 10/20, not 50%)

Outcomes and estimation	16	For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g. 95% CI)	a) Inclusion of null and negative findings b) Inclusion of results from testing pre-specified causal pathways through which the intervention was intended to operate, if any
-------------------------	----	-------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Ancillary analyses	17	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating which are pre-specified and which are exploratory
--------------------	----	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Although the study was not powered for this hypothesis, exploratory an exploratory analysis shows that the intervention effect was greater among women than among men (although not statistically significant)

Adverse events	18	All important adverse events or side effects in each intervention group	Events which affected either the treatment or control groups which threatened the integrity of the design.
----------------	----	-------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------

'Police cracked down on prostitution, which drove the target population, commercial sex workers, to areas outside the recruitment/sampling area' or 'Midway through the trial, an international NGO began implementation of a similar program in a significant proportion of the study control areas.'

Discussion

Interpretation	19	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes	<p>a) Discussion of results taking into account the mechanism by which the intervention was intended to work (causal pathways) or alternative mechanisms or explanations</p> <p>b) Discussion of the success of and barriers to implementing the intervention, including acceptability and fidelity of implementation</p> <p>c) Discussion of research, programmatic, and/or potential policy impact/implications based on cost effectiveness, or cost-benefit, analysis wherever appropriate</p>	
Generalizability	20	Generalisability (external validity) of the trial findings	Generalizability (external validity) of the evaluation findings, taking into account the study population, the characteristics of the intervention, length of follow-up, settings and other contextual issues.	<p>'An important limitation of large scale social experiments, such as PROGRESA, is that it is often prohibitively costly to vary the experimental treatments in a way that permits evaluation of a variety of policies of interest. In the PROGRESA experiment, all eligible treatment group households faced the same subsidy schedule, so it is not possible to evaluate the effects of alternative subsidy schemes through simple comparisons of treatments and controls. In addition, because the experiment lasted only two years, one cannot directly assess the long term impacts of the program on completed schooling, or ... compare the effects of the existing subsidy program to the effects of various alternative (non-existent) programs' (Todd and Wolpin, 2006 as described in Duflo et al 2008).</p>

References

3ie (2010) "Impact Evaluation Glossary" *Journal of Development Effectiveness* 2(1), also available at www.3ieimpact.org.

Bruhn, Miriam & McKenzie, David, 2008. "In pursuit of balance : randomization in practice in development field experiments," Policy Research Working Paper Series 4752, The World Bank.

Blattman, Chris, 2008 "Impact Evaluation 2.0" DFID presentation, Accessed September 22, 2009, available at:
http://www.chrisblattman.com/documents/policy/2008.ImpactEvaluation2.DFID_talk.pdf

Cohen, J. and P. Dupas (2007) 'Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment'. *Brookings Global Economy and Development Working Paper No. 11*. Available at SSRN: <http://ssrn.com/abstract=1080301>

De Mel, Suresh, David McKenzie, and Christopher Woodruff, 2009, "Are women more credit constrained?," *American Economic Journal: Applied Economics*, Volume 1 Issue 3.

Duflo, E., R. Glennerster, and M. Kremer, 2008, "Using Randomization in Development Economics Research: A Toolkit," *Handbook of Development Economics*, (T. P. Schultz and J. Strauss eds.), 3895-3962.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2009) 'Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya', available at: http://www.povertyactionlab.org/papers/61_Duflo_Peer_Effects_and_Tracking.pdf

Fernald, Lia Rita Hamad, Emily J. Ozer and Jonathan Zinman., 2008, "Small individual loans and mental health: a randomized controlled trial among South African adults", *BMC Public Health*, volume 8: 409.

Karlan, Dean and Martin Valdivia (2009) 'Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions', New York: Financial Access Initiative, and New Haven: Innovations for Poverty Action.

Moher D, Schulz KF, Altman DG, 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 357:1191–1194., Accessed September 22, 2009, available at: www.consort-statement.org

REAP (2009) 'Anemia in Rural China's Elementary Schools: Prevalence, Correlates and Consequences in Shaanxi Province's Poor Counties', REAP Working Paper #211, 2009 available at:
http://iis-db.stanford.edu/pubs/22724/211_Anemia_in_Rural_Chinas_Elementary_Schools.pdf

White, Howard, 2009a, "Some reflections on Current Debates in impact evaluation", 3ie Working paper 1, available at: http://ideas.repec.org/p/ris/iiierp/2009_001.html (forthcoming in *Evaluation*).

White, Howard, 2009b, "Theory-Based Impact Evaluation: principles and practice", *Journal of Development Effectiveness* 3(1): 271-284.

Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D. 2008 "Improving the reporting of pragmatic trials: an extension of the CONSORT statement".

Notes

ⁱ See White (2009) for the debate on the definition of impact evaluation (endorsed in the recently released NONIE guidance) by the DAC committee is "the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. These effects can be economic, socio-cultural, institutional, environmental, technological or of other types". By contrast the definition implied in most recent impact evaluations, including RCTs, refers to rigorous attribution, i.e. by identifying the (unobserved) counterfactual to calculate the outcome in the absence of the intervention (White, *ibid*).

ⁱⁱ See the 3ie impact evaluation glossary for an elaboration of terms (3ie, 2010).

3ie Working Papers

Impact evaluation in the post-disaster setting by Alison Buttenheim, 3ie Working Paper 5, December 2009

Designing impact evaluations: different perspectives, contributions from Robert Chambers, Dean Karlan, Martin Ravallion, and Patricia Rogers, 3ie Working Paper 4, July 2009

Theory-Based Impact Evaluation by Howard White, 3ie Working Paper 3, June 2009

Better Evidence for a Better World edited by Mark W. Lipsey University and Eamonn Noonan, 3ie & The Campbell Collaboration, 3ie Working Paper 2, April 2009

Some Reflections On Current Debates In Impact Evaluation by Howard White, 3ie Working Paper 1, April 2009