

Pseudo-Placebo Effects in Randomized Controlled Trials for Development: Evidence from a Double-Blind Field Experiment in Tanzania

Erwin Bulte,^{1*} Lei Pan,¹ Joseph Hella,² Gonne Beekman¹ and Salvatore di Falco³

1: Development Economics Group, Wageningen University, P.O. Box 8130, 6700 EW
Wageningen, Netherlands

2: Sokoine University of Agriculture, Morogoro, Tanzania

3: London School of Economics, London, United Kingdom

(*Corresponding author: Erwin.bulte@wur.nl)

Abstract: Randomized controlled trials (RCTs) in the social sciences are not double-blind, so participants know they are “treated” and will adjust their behavior accordingly. Under some conditions this gives rise to so-called “pseudo-placebo effects,” which may bias assessment of impact. We implement a conventional economic RCT and a double-blind experiment in rural Tanzania (randomly allocating modern and traditional cowpea seed-varieties to farmers), and demonstrate that such pseudo-placebo effects can be large. For our case they explain the entire “treatment effect on the treated” as measured in a conventional economic RCT.

Keywords: Randomized controlled trial (RCT), behavioral response, experimenter effect, placebo

JEL Codes: C9, D04,

1. Introduction

Randomized controlled trials (RCTs) have transformed the economic landscape in recent years, and are a key component of the “credibility revolution” that has reinvigorated empirical economics (e.g. Angrist and Pischke 2010). Random assignment of units to treatment or control group ensures exogeneity of key variables, so that a straightforward comparison of sample means yields reliable estimates of average treatment effects (ATE). RCTs are applied in a variety of domains to evaluate the impact of a wide range of development interventions. Recent examples include interventions in the domains of health, education, microfinance, food production, technology adoption and institutional reform (Banerjee and Duflo 2011).

The RCT revolution in empirical economics has not gone unopposed. Among the concerns are various statistical issues (e.g. about non-compliance and spillover benefits), aggregation issues (e.g., external validity, general equilibrium effects), ethical issues, and practical issues (randomization may not be feasible or cost-effective for many relevant interventions—see Rodrik 2008, Deaton 2010). Others emphasize conceptual and statistical difficulties due to heterogeneity among the sample population (Ravallion 2011, Harrison 2011), or remind us that policy makers are ultimately not interested in treatment effects but in welfare changes (so that structural modeling should complement RCTs—see Harrison 2011). While these considerations, and others, are well known, experimenters claim that RCTs offer more reliable evidence on causation than observational studies. Imbens (2011) concludes “randomized experiments do occupy a special place in the hierarchy of evidence, namely at the very top.” Analysts sometimes refer to RCTs as the Gold Standard in impact evaluation.

In this paper we highlight a novel concern that may confound impact assessment via RCTs. While proponents of the RCT approach have drawn parallels between their work and

that of analysts in medical science, there is a significant difference between medical and economic trials: the (genuine) gold standard prescribes *double-blind* implementation of trials. Patients in the control group receive a placebo, and neither researchers nor patients know the treatment status of individuals. The reason is that patients' perceptions and expectations have therapeutic effects. Even in the absence of active ingredients (i.e. when using inert or sham drugs), interventions may produce a subjective perception of medical improvement, and even improve the physical condition of participants. The strength of placebo effects varies across social groups and conditions, depending on the degree of belief in beneficial outcomes (Malani 2006). Failing to control for placebo effects implies overestimating the impact of the intervention.

Such placebo effects may also be relevant for RCTs in the domain of economic development, where double-blind interventions are *not* the standard. We do not introduce sham microfinance groups or fake clinics as the “social science counterpart” of inert drugs when analyzing the impact of interventions in the credit or health domain—respondents *know* when they are “treated.” Placebo effects may materialize when subjective welfare criteria are used to assess the impact of any intervention—respondents could feel better just because they participate in a program, believing their concerns have been taken seriously and addressed. This resulting gain in subjective welfare, however, should not be attributed to a *specific* intervention.

A variant of the placebo effect materializes if an intervention affects the value marginal product of (other) inputs provided by the participant, because then treatment status invites a behavioral response. The household production model emphasizes that participants typically combine a range of inputs—privately provided and otherwise—to produce the outcomes they desire, including health status, education, food security, etc. While random assignment ensures that the intervention is orthogonal to *ex ante* participant characteristics,

treatment and control groups will be different *ex post* because treated individuals behave differently. We will demonstrate that distinguishing between treatment and behavioral effects may not be straightforward, and impact assessments could be biased in unknown directions.

“Randomistas” may find such reasoning irrelevant. They will argue that, from a policy perspective, we are exactly interested in the total effect of an intervention (i.e. the sum of the intervention effect *per se* and the optimal behavioral response to the new realities), and not in the outcomes of a double-blind experiment. We do not disagree. But the second order attribution problem—*are outcomes as measured in an RCT due to the intervention or a behavioral response?*—makes explicit that impact estimates of conventional RCTs may be biased for at least two reasons. First, insofar as the analyst fails to control for all relevant behavioral adjustments (there may be many dimensions along which behavior can be adjusted!) the eventuating impact assessment is biased. This is not a new observation. For example, Duflo et al. (2008) seek to assess the rate of return on fertilizers, and correctly highlight the importance of measuring the impact “on the use of complementary inputs” as well as on output. But in practice it is often very difficult to pick up changes in the use of all complementary inputs.¹

Second, even if analysts are able to monitor adjustments across the full range of complementary inputs, then the eventuating results would only provide an unbiased assessment of the total impact (sum of intervention and optimal response) if the participants have sufficient information to optimally adjust their input allocation. But this creates a

¹ For example, Duflo et al (2008) focus on differences between treatment and control plots in the time that farmers spent weeding, and on enumerators’ observations of the physical appearance of the plot. They detect no differences and therefore assume that “costs other than fertilizer were similar between treatment and control plots” (p.484). This may be true, but it is also possible that these analysts have underestimated the complexity of the farm household system and the associated heterogeneity in production conditions at the village or farm level. Our empirical analysis below illustrates this point.

paradox. We organize an RCT to learn about impact, so to assume that farmers correctly anticipate the nature of the innovation that is offered to them, and will optimally adjust the provision of complementary inputs, is typically erroneous. Instead, the behavioral response picks up subjective beliefs of participants, and possibly risk preferences. In what follows we propose to dissect the total impact of an intervention in the innovation effect (given a certain level of complementary inputs) and a “*pseudo-placebo effect*,” which captures the behavioral response to intervention status. This pseudo-placebo effect captures changes in behavior due to the beliefs and expectations of participants, which may or may not be accurate.

In this paper we use experimental evidence from an agricultural development intervention in central Tanzania to test whether thus-defined pseudo-placebo effects may be relevant when evaluating the impact of development interventions. We distributed modern and traditional seed-varieties among random subsamples of farmers, and compared the outcomes of a double-blind RCT with the outcomes of a conventional economic RCT (a dif-in-dif-in-dif model). We find that all impact that would routinely be attributed to the modern seed intervention is, in fact, due to a pseudo-placebo effect. Farmers who were unsure about the quality of their seed (i.e. farmers in the double-blind experiment) and farmers who knew they received the modern seed chose to plant their seeds further apart than farmers who knew they received traditional seed in the conventional economic RCT (control group). This response increased harvests, but also involved costs as it implied reduced opportunities to grow other crops. This re-alignment of complimentary inputs, however, was not obvious (at least: not to us, certainly not *ex ante*) and we speculate it would easily be overlooked. Moreover, our results indicate that the expectations and beliefs of the participating farmers with respect to the productivity of the modern cowpea seed were overly optimistic. A one-shot RCT would measure impact even if the treatment is ineffective—we inadvertently introduced a sham innovation.

This paper is organized as follows. In section 2 we introduce and explain the pseudo-placebo effect. In section 3 we describe our experiments, data, and identification strategy. Section 4 contains our results. We demonstrate that the difference between treatment and control group in a conventional RCT is due entirely to a pseudo-placebo effect, and are able to identify the mechanism via which farmers manage to raise harvests (the spacing of seeds). In section 5 we speculate about the implications for policy makers, who ultimately are interested in the *net* effect of an intervention (capturing both the intervention and the behavioral response).

2. Introducing the pseudo-placebo effect

The experimental literature identifies various effects, in addition to the above-mentioned placebo effect, that may preclude causal inference when experiments are not double-blind. These include the Pygmalion effect (expectations placed upon respondents affecting outcomes) and the observer-expectancy effect (cognitive bias unconsciously influencing participants in the experiment). Behavioral responses may also originate at the respondent side. Well-known examples are the Hawthorne effect (capturing that respondents in the treatment group change their behavior in response to the fact that they are studied—see Levitt and List 2011) and the opposing John Henry effect (which captures bias introduced by reactive behavior of the control group). Relatedly, Zwane et al. (2011) demonstrate the existence of so-called “survey effects” (i.e., being surveyed may change later behavior).

In addition to these potentially confounding effects, optimizing participants should adjust their behavior if an intervention affects the relative returns of their behavior. Consider a population of farmers. We know since Schultz (1964) that smallholder farmers are rational optimizers, who respond to new opportunities. Assume each farmer combines a private input l and seed x to produce a crop: $y = f(l;x)$. Also assume $f_l > 0$ and $f_{ll} < 0$, and that there are two

varieties of seed, $x=\{M,T\}$ where $x=M$ represents a modern, improved variety and $x=T$ represents a traditional, less productive seed type. In what follows we assume $f(l^*;M) > f(l^*;T)$ and $f_{l^*|M} > f_{l^*|T}$, $\forall l^* \in [0,L]$ where L is the endowment of the private input. The farmer maximizes utility, $U = y - c(l)$, where c captures the cost of the input ($c' > 0$, $c'' > 0$), so that the optimal amount of input allocated to cropping solves $f_{l|x} = c'$.

Next, we distinguish between two experimenters interested in the welfare effects of distributing the modern seed variety. The first experimenter adopts a conventional economic RCT so that farmers know whether they receive the modern or traditional seed. Farmers in the intervention group solve $f_{l|M} = c'$, and supply \bar{l} units of input on the farm. Farmers in the control group solve $f_{l|T} = c'$, and supply only \tilde{l} units. The second experimenter implements a double-blind experiment with other sub-samples of farmers, and the farmers in this experiment will supply \hat{l} units, so that $f_{l|x=\{M,T\}} = c'$. Since seed quality raises the marginal product of the input, the following holds: $\bar{l} > \hat{l} > \tilde{l}$.

Assuming that the first experimenter focuses on observables (harvest), she obtains as a measure of the average treatment effect: $\bar{y}(\bar{l}, M) - \tilde{y}(\tilde{l}, T)$. This expression confounds the positive effects of modern seeds and the extra input put in by the farmer, and *overestimates* the true welfare effect ($\Delta U = \bar{U} - \tilde{U}$) associated with the distribution of modern seed. The reason, of course, is that cost associated with the private input has increased as well: $c(\bar{l}) > c(\tilde{l})$.² While estimates of the treatment effect in this example may be “corrected” by closely monitoring the allocation of inputs ($\bar{y}(\bar{l}, M) - \tilde{y}(\tilde{l}, T) + c(\tilde{l}) - c(\bar{l})$), such corrections are clearly impractical in real life situations where harvests are potentially determined by the

² The welfare effect would be underestimated if, instead, the intervention amounts to promoting an input-saving technology so that $\bar{l} < \tilde{l}$.

interplay between a vector of inputs (including plot size and location, soil quality, quantity and quality of labor, fertilizer, irrigation, etc.).

The average treatment effect according to the second experimenter is $\hat{y}(\hat{l}, M) - \hat{y}(\hat{l}, T)$. Call this treatment effect the seed effect. Due to the double-blind nature of her experiment, the seed effect is not confounded by changes in the supply of inputs (as $l = \hat{l}$ for both sub-groups) and allows clean identification of the contribution of modern seed (for $l = \hat{l}$). However, since this experiment does not enable the farmer to optimally adjust his input allocation to the opportunities provided by the modern seed, this treatment effect *underestimates* the true potential welfare gains associated with the distribution of modern seeds.

We obtain a new variable if we combine the data of the two experimenters. Define $\hat{y}(\hat{l}, T) - \tilde{y}(\tilde{l}, T)$ as the *pseudo-placebo effect* associated with seed distribution. It captures extra output not due to modern seeds per se, but due to the belief that the seed may be of the modern variety. Consistent with the conventional placebo effect, it represents a system response to expectations about the intervention. Unlike the conventional placebo effect, the relevant system is not the human body but a farm household. Hence the response does not follow from an unconscious physiological reaction—it reflects the purposeful reallocation of inputs such that expected profits, conditional on expectations and beliefs, are maximized. The magnitude of the pseudo-placebo effect is increasing in expectations and beliefs regarding the complementarity of labor and modern seed type (regardless of whether such beliefs are accurate or not, see Malani 2006). Pseudo-placebo effects can dominate measures of the innovation effect. Indeed, they could be all there is.

3. Data and identification

We conducted two experiments in Morogoro Region in Tanzania in February-August 2011. Eight hamlets³ in the vicinity of Morogoro city were selected, which together comprises a village (Mikele). They are located along a road connecting Dar es Salaam to Zambia and the Democratic Republic of Congo. Main income activities are agriculture and trade. Farm households typically crop multiple plots, which is common in Africa. From the eight hamlets, nearly 600 household representatives were randomly selected to participate in the experiment, and randomly allocated to one of four treatment-groups.

Consistent with the simple model above, we did two experiments: (i) a conventional economic RCT and (ii) a ‘double-blind’ RCT. Participants in the conventional economic RCT received cowpeas of either the modern type (group 1) or the traditional type (group 2), and the experimenter informed the participants about the type of seed they received. The participants in the double-blind RCT received cowpeas of either the traditional type (group 3), or the modern type (group 4), but neither the experimenter nor the participants were aware which type of seed was provided. All seed was distributed in closed, paper bags, and the seed-types were indistinguishable from one another, both in terms of size and color.⁴

All participants were informed that upon accepting the seed, they could plant the seed but were not allowed to mix it with their own cowpeas. They were also informed that the yield fully belonged to them. Each participant received a bag to separately store the

³ The participating hamlets were: Dindili, Lukole, Mtego wa Simba, New Land A, New Land B, Nungu, Ukomanga, and Ulundo.

⁴ For the objective of our study it was crucial that the traditional and modern seed looked exactly the same. In the double-blind RCT, neither experimenters nor participants should be able to identify the actual seed-type at first sight (even if information about the type may be gradually revealed as the crop matures in the field, but then key inputs have already been provided). It turned out, however, that modern seed was treated with purple powder, to prevent cheating on the market place and to protect the seed from insect damage during storage. We hence treated the traditional seed with the same purple powder. Our strategy worked, because when the participants received the seed, 96% of the participants in the double-blind RCT did not know what seed-type they received, and of the remaining participants, half of the people guessed the seed-type wrong. In contrast, nearly all participants in the conventional economic RCT knew which seed-type they had received.

harvested cowpeas until an experimenter had measured the yield. Seed was planted during the onset of the rainy season (February-March), and harvested when the dry season started (June-July).

After the seed was accepted, the experimenter conducted a household survey which included sections about demographic characteristics, welfare, land use, plot characteristics, cowpea planting techniques, labor allocation, income activities, and consumption. By the end of the growing season, the experimenters conducted field measurements on all plots where the participants had planted their cowpeas. Field measurements included detailed measurement of plot size, number of plants grown, number of pods per plant, and observations concerning quality of land (slope, erosion, weeding). After harvest, the experimenters visited the participants to weigh the total harvest and to conduct an endline survey. The endline survey included questions about beliefs about the type of seed, the use of the distributed cowpeas, and the resulting yield.

We use two dependent variables. First, our preferred variable is the total harvest of cowpeas. Cowpeas are harvested on an on-going basis towards the end of the growing season, and we asked farmers to store their harvests in the special bag we provided. Our harvest variable thus captures harvests during the cropping cycle as well as the final harvest. Second, and as a robustness analysis, we also did a field measurement towards the end of the cropping cycle, counting the number of remaining pods in the field. This variable proxies for total production with quite some noise, as it does not capture prior harvests during the cropping cycle (which vary from one farmer to the next, depending on a range of variables). Hence, our pod count variable is a more noisy signal of total production than our harvest variable (as is also clear from the summarized data in Table 2), and this will attenuate the statistical significance of our findings. Insofar as on-going harvesting varies with seed type, the pod count variable will also be biased.

Finally, attrition in our sample is large—over 40% of the initial sample did not harvest cowpeas from the seeds we provided them with. The main reasons are failed harvests (water damage, insect attacks) and failure to plant the distributed seeds. Attrition, however, was rather equally spread across our 4 groups, and did not seem to introduce selection bias. ANOVA tests for all 44 household characteristics collected at the baseline shows that we cannot reject the null hypothesis of no difference between the four treatment groups for no less than 43 variables (the exception being the percentage of household members younger than age 6). Table 1 reports a selection of these variables, and associated P-values of the ANOVA test.

4. Results

Table 2 contains our main results, summarizing harvest and pod-count data for the 4 different groups. We focus the discussion on our preferred harvest variable (the top row), and note that patterns in the data for the pod count variable are similar, but less significant (due to greater dispersion of these data, as expected). Columns 1 and 2 present the outcomes of the conventional RCT. The harvest variables for the modern seed type are 27% greater than for the traditional seed type, and t-tests confirm this difference is statistically significant at the 5% level. A naïve analyst would interpret this as evidence that modern seeds raise farm output. Based on such an interpretation, policy makers could consider implementing an intervention that consists of distributing modern seeds to raise rural income or improve local food security (depending on the outcomes of a complementary cost-benefit analysis, hopefully).

A different picture emerges when we look at the outcomes of the double-blind experiment, summarized in columns 3 and 4. When farmers are unaware of the type of seed allocated to them, the modern type does not outperform the traditional type (given the particular farmers' response in terms of complementary inputs). The seed effect, defined

above, is zero. Based on evidence of this double-blind experiment, an economic analyst is unlikely to recommend large-scale implementation of a modern seed intervention.

Additional insights emerge when we combine the evidence from the RCT and double-blind experiment. First, comparing the performance of the modern type across experiments (groups 1 and 4), we find that outcomes are similar. The most interesting comparison, however, is the comparison of groups 2 and 3—output for the traditional seed-type varies significantly depending on the information available to farmers. The difference between groups 2 and 3 is our measure of the pseudo-placebo effect: it captures the crop and harvest differential due to beliefs and associated behavioral responses, not to modern seed. In the first row, which captures our preferred dependent variable (total harvest), the pseudo-placebo effect explains the entire difference between seed-types as picked up in a standard RCT. The pseudo-placebo effect is 23%, and group 3 is statistically indistinguishable from group 1.⁵

Why are harvest and pod-count lower when farmers are in the control group of the standard RCT? We probe into this question in Table 3, which compares key inputs and conditioning variables across the three groups of farmers (that is: the two RCT groups, and the aggregate double-blind group). The ANOVA test suggests differences in terms of soil quality and plot size. Pairwise comparisons of the groups reveal that (i) farmers in the RCT receiving the modern seed chose to plant this seed on good quality plots, and (ii) farmers receiving traditional seeds in the RCT chose to plant on relatively small plots. While differences in soil quality do not translate into significantly different harvests (Table 2), it appears as if decisions with respect to plant density matter for harvest. Spacing affects output. However, and compensating for the lower cowpea harvest, farmers in group 2 could allocate

⁵ An interesting question is why a pseudo-placebo effect materializes for traditional seed but not for modern seed. A priori we would expect that uncertainty about treatment status would invite a relative “under-supply” of inputs for the modern seed in the double-blind experiment. Indeed, a closer look at the data suggests that on average farmers in group 1 allocate better quality plots to their seed than farmers in the double-blind experiment (groups 3 and 4, see Table 3). However, this does not translate into significantly greater harvests.

their larger plots to other crops, and presumably outperformed their peers along this other farm household activity.

Two important observations for impact evaluation stand out. First, it may be difficult to capture all relevant adjustments in complementary inputs. Farmers can optimize along multiple dimensions, and failing to control for all of them will result in biased estimates of impact. A priori we did not expect that plot size would be the crucial mechanism (indeed: we cannot be certain that we have not overlooked other relevant mechanisms!). Failing to control for the plot size mechanism implies overestimating the impact of modern seed.

Second, farmers had overly optimistic beliefs about the productivity of modern seed. Farmers were disappointed by the eventuating harvests—nearly 60% of the farmers receiving the “modern variety” indicated to have lower production levels than the year before (57% in group 1, 59% in group 4). Since modern seed does not outperform traditional seed, and since farmers prefer to plant traditional seed on smaller plots, we conclude the farmers have allocated too much land to modern cowpeas. This comes either at the cost of other crops or, if the alternative land use would have been to fallow it, it could come at the expense of soil fertility and future productivity.⁶ If we would repeat the experiment, farmers would presumably revert back to smaller plots. Failing to control for this introduces another source of bias.

Finally, we discuss another reason why the pseudo-placebo effect is relevant. Assume an analyst uses randomization into treatment as an instrumental variable to compute the Local Average Treatment Effect (LATE). When all participants accept their assignment, farm output Y is determined by $Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 L_{1,i} + \alpha_3 L_{2,i} D_i + \varepsilon_i$ where D indicates assignment

⁶ In the case of cowpeas, the effect on soil fertility might be positive, given the nitrogen-fixing nature of peas. Pea varieties are often used as alternative fertilizer on otherwise fallow land. Reduced fallowing would, however, have negative effects on soil fertility in the case of most other crops.

to the treated group (use of modern seeds), L_1 denotes land used in case of traditional seeds, L_2 is extra land for modern seeds, and ε is the error term. If we use the difference between the means of the outputs as an estimate of the program effect, we can estimate the following equation using OLS: $Y_i = \alpha + \beta D_i + v_i$. The expectation of the OLS estimator of β is $E(\beta) = \alpha_1 + \alpha_3 E(L_2)$ where E denotes the expectations operator. Hence, as before, β sums the seed effect α_1 and the average pseudo-placebo effect $\alpha_3 E(L_2)$. To disentangle these effects, the analyst needs to observe the variable L_2 .

If the take-up of assignment status is voluntary (so that not all respondents comply with their role), program output is determined by: $Y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 L_{1,i} + \alpha_3 L_{2,i} T_i + \varepsilon_i$ where T denotes voluntary take up. The usual practice is to use program assignment D as an instrumental variable (IV) for program treatment T , or estimating: $Y_i = \alpha + \beta D_i + u_i$ using a two-stage model (where D is the IV of T). The expectation of the IV estimator of β is $E(\beta) = \alpha_1 + \alpha_3 E_{T=1}(L_2)$, where $E_{T=1}(L_2)$ denotes the expectation of L_2 for households with $T=1$. The IV estimator sums the seed effect and the mean pseudo-placebo effect of the treated, but does not provide an unbiased estimate of the overall mean effect. Heterogeneity of the program effect—in our case heterogeneity of the pseudo-placebo effect, driven by diverging expectations or beliefs—renders the use of program assignment D as the IV for program take-up T invalid (also see Ravallion 2011).

5. Implications and conclusions

Conventional economic RCTs do not meet the requirements of the gold standard as applied elsewhere in science because they are not double-blind. Participants know whether they are treated or not, and adjust their behavior accordingly. While random assignment implies that intervention status is exogenous to participant characteristics, it generally is not orthogonal to behavior. This introduces a confounding effect that we label as the “pseudo-

placebo effect.” Our case study suggests that the pseudo-placebo effect can be large—accounting for the full difference between treated and control group in a conventional economic RCT. A naïve analyst would have concluded that modern cowpea seed raises output. The double-blind analysis learns us that the increment in output is not due to the quality of the seeds at all, but instead to a behavioral response of the farmer (who plants the – suspected – modern seeds further apart). Positive treatment effects evaporate when controlling for this. The main conclusion is not to shun RCTs—far from it. However, we should interpret the outcomes of RCTs with proper caution, especially when jumping from output to outcomes and impact.

The magnitude of the pseudo-placebo effect depends on the participants’ subjective expectations with respect to the degree of complementarity of the intervention and privately supplied inputs. Biased impact assessments result when the analyst is unable to capture all relevant adjustments. The combination of conventional RCTs and double-blind experiments provides relevant information about which behavioral adjustments occur. Double-blind experiments are also helpful to evaluate whether or not the innovation has any impact that is not due to differences in complementary inputs.

Biased assessments also occur when participants have the wrong set of expectations about complementarities. To attenuate this source of bias it would be advisable to repeat experiments over various production cycles, allowing participants to learn about the nature of the innovation offered to them. As knowledge accumulates, their response becomes better tailored to the innovation, enabling the analyst to obtain an estimate of the total impact of interventions that captures both the innovation effect as well as the optimized behavioral response. Ultimately this is the information that will be valuable to policy makers.

References

- Angrist, J. and J-S Pischke, 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24: 3-30
- Banerjee, A. and E. Duflo, 2011. *Poor economics: A radical rethinking of the way to fight global poverty*. New York: PublicAffairs
- Deaton, A., 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* 48: 424-455
- Duflo, Esther C., Michael R. Kremer and Jonathan M. Robinson. 2008. "How High are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review Papers (Papers and Proceedings Issue)*, 98 (2): 482–488.
- Harrison, G., 2011. Randomisation and its discontents. *Journal of African Economies* 20: 626-652
- Imbens, G., 2010. Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature* 48: 399-423
- Levitt, S. and J. List, 2011. Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics* 3: 224-238
- Malani, A., 2006. Identifying Placebo Effects with Data from Clinical Trials. *Journal of Political Economy* 114: 236-256.

Ravallion, M., 2011. On the implications of essential heterogeneity for estimating causal impacts using social experiments. Policy Research Working Paper 5804, Washington DC: World Bank

Rodrik, D., 2008. The new development economics: We shall experiment, but how shall we learn? Working Paper 2008-0142, Weatherhead Center for International Affairs, Harvard University

Schultz, T., 1964. *Transforming traditional agriculture*. New Haven: Yale University Press

World Bank, 2007. *World Development Report 2008: Agriculture for Development*. Washington D.C.

Zwane, A.P., J. Zinman, E. van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. Gine, E. Duflo, F. Devoto, B. Crepon and A. Banerjee, 2011. Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences* 108: 1821-1826

Table 1: Did Randomization “work”? A Sample of Observables for the Four Groups

Variables	<i>Economic RCT</i>		<i>Double-blind</i>		ANOVA test (P-value)
	Improved	Traditional	Improved	Traditional	
	Group 1 (N=70)	Group 2 (N=73)	Group 3 (N=63)	Group 4 (N=56)	
Household size	4.886 (2.646)	4.959 (2.786)	5.048 (2.043)	4.875 (2.141)	0.97
Gender household head (1=male)	0.692 (0.465)	0.794 (0.408)	0.754 (0.434)	0.736 (0.445)	0.60
Years of education household head	2.625 (3.068)	2.516 (3.050)	2.933 (3.732)	2.654 (3.093)	0.90
Age household head	48.339 (16.470)	49.778 (15.391)	47.169 (16.108)	50.769 (16.971)	0.65
Dependency ratio	0.523 (0.270)	0.543 (0.283)	0.541 (0.240)	0.519 (0.258)	0.94
Village leaders' household and their relatives (1=yes)	0.186 (0.392)	0.178 (0.385)	0.111 (0.317)	0.286 (0.456)	0.11
Members of economic groups (1=yes)	0.243 (0.432)	0.233 (0.426)	0.206 (0.408)	0.143 (0.353)	0.53
Members of social groups (1=yes)	0.357 (0.483)	0.384 (0.490)	0.333 (0.475)	0.268 (0.447)	0.57
Health (1=somewhat good or good)	0.515 (0.503)	0.458 (0.502)	0.550 (0.502)	0.509 (0.505)	0.77
Economic situation compared to village average (1=somewhat rich or rich)	0.294 (0.459)	0.306 (0.464)	0.295 (0.460)	0.255 (0.440)	0.93
Land owned (acre)	3.880 (5.664)	4.865 (6.044)	4.921 (6.168)	4.410 (4.916)	0.72
Own a bike (1=yes)	0.457 (0.502)	0.438 (0.500)	0.444 (0.501)	0.339 (0.478)	0.55
Value of productive assets (1000 Tsh)*	6.644 (7.060)	7.285 (7.642)	7.388 (6.645)	7.598 (7.843)	0.89
Value of other assets (1000 Tsh)*	155 (250)	136 (194)	149 (225)	176 (301)	0.83
Food consumption 7 days (1000 Tsh)*	29.468 (7.498)	31.778 (7.843)	29.969 (8.646)	32.124 (9.289)	0.19

*Observations in the top 3 percentiles of the variable are dropped when calculating the mean and the standard deviation.

Table 2: Treatment and Pseudo-Placebo Effects: Dependent Variables for the 4 Groups

Variables	Economic RCT		Double-Blind		P-value of t-test			
	Improved Group 1	Traditional Group 2	Traditional Group 3	Improved Group 4	Group 1=2	Group 3=4	Group 1=4	Group 2=3
Total harvest in seeds (kg)	9.865 (10.809)	7.238 (6.175)	9.400 (8.614)	9.912 (10.012)	0.05	0.72	0.97	0.06
Pods during endline field measurement*	17,806 (29,217)	12,285 (13,048)	16,270 (15,122)	14,997 (22,017)	0.14	0.71	0.55	0.10

*Observations in the top 3 percentiles of the variable are dropped when calculating the mean and the standard deviation.

Table 3: What Explains Higher Harvests?

Variables	Group 1	Group 2	Group 3/4	ANOVA test (P-value)	P-value of t-test		
					Group 1=2	Group 1=3/4	Group 2=3/4
Household labour on cowpea	9.273 (5.789)	10.354 (8.039)	9.654 (6.749)	0.59	0.33	0.67	0.47
Land is flat (1=yes)	0.319 (0.469)	0.421 (0.497)	0.369 (0.484)	0.44	0.20	0.48	0.46
Land erosion (1=slight or heavy erosion)	0.712 (0.456)	0.632 (0.486)	0.699 (0.461)	0.50	0.29	0.83	0.32
Improvement such as bounding, terrace (1=yes)	0.263 (0.443)	0.244 (0.432)	0.242 (0.430)	0.94	0.78	0.73	0.97
Intercropping (1=yes)	0.186 (0.392)	0.159 (0.369)	0.146 (0.355)	0.77	0.68	0.47	0.80
Weed between plants (1=yes)	0.819 (0.387)	0.681 (0.470)	0.746 (0.437)	0.15	0.05	0.23	0.32
Soil quality (1=good)	0.671 (0.473)	0.471 (0.502)	0.476 (0.501)	0.01	0.01	0.01	0.94
Consult anybody on how to plant cowpea (1=yes)	0.139 (0.348)	0.118 (0.310)	0.158 (0.366)	0.53	0.51	0.70	0.26
Plot size (square meter)*	351 (224)	293 (220)	366 (235)	0.10	0.13	0.68	0.04

* Observations in the top 3 percentiles of the variable are dropped when calculating the mean and the standard deviation.