A New Perspective on the Issue of Selection Bias into Randomized Controlled Field Experiments

Michèle Belot
University of Edinburgh

Jonathan James
University of Bath

June 4, 2014

Abstract

Many randomized controlled trials require participants to opt in. Such self-selection could introduce a potential bias, because only the most optimistic may participate. We revisit this prediction. We argue that in many situations, the experimental intervention is competing with alternative interventions participants could conduct themselves outside the experiment. Since participants have a chance of being assigned to the control group, participating has a direct opportunity cost, which is likely to be higher for optimists. We propose a model of self-selection and show that both pessimists and optimists may opt out of the experiment, leading to an ambiguous selection bias.

Keywords: Field experiments, selection bias, randomized controlled trials, external validity
JEL codes: C4, C9

1. Introduction

The last decade has seen a booming increase in popularity of field experiments in economics (Holt, 2005) and there is now a stronghold of researchers advocating the case for randomized controlled trials (RCTs) to overcome the challenges of endogeneity (Burtless, 1995, Duflo and Kremer, 2005). One key issue is external validity, specifically, we would like to know if we can generalize the lessons we draw from such experiments to the wider population.[1] The answer to this question depends on the representativeness of the sample along characteristics that may be correlated with the estimated treatment effects. In that context, self-selection deserves prime attention.

[1] A number of recent papers study self-selection in field experiments: Allcott and Mullainathan (2012), Belot and James (2013), Gautier and van der Klaauw (2012).

In almost all randomized controlled experiments in social sciences, participation is voluntary. Researchers are well aware of the implications of self-selection for external validity. The standard prediction in the literature on selection in randomised field experiments is a positive selection bias, following the spirit of the literature on policy evaluation (Heckman and Vytlacil, 2006 and Heckman et al., 1999). List and Rasul (2011) state that "*Indeed, in almost any social experiment related to job training programs, it is a concern that those most likely to benefit from the program select into the program.*"

We argue that this prediction is not necessarily correct because participating to the experiment does not guarantee treatment and entails a chance of ending up in the control group. Being assigned to the control group often entails an opportunity cost. Participants are often required not to conduct any competing intervention (and certainly not the experimental intervention itself) at the same time. Almost all RCTs (in medical and social sciences) entail such opportunity cost.

In many situations, the intervention to be tested already exists and the main goal of the RCT is to establish causality. The key reason why causality is difficult to establish is because of self-selection: Those who are already exposed to the treatment are not a random sample of the population and are likely to be those who benefit most from the treatment. In fact, researchers often look for "virgin samples", i.e. participants who have not yet been exposed to the treatment. Or, alternatively, participants themselves will not see the point of participating if they are already implementing the intervention. For example, a firm that implements a tournament incentive scheme may not be interested in testing the effects of a tournament, particularly if it means there is a chance they may have to give it up temporarily (or have a smaller proportion of their workers exposed to the tournament scheme) if they end up in the control group. Of course this obviously introduces a selection bias which, surprisingly, has received little attention in the literature.

2. The Model

A. The traditional problem of selection bias

Consider, the average treatment effect, that is the treatment effect averaged over the population of interest. In the case of policy evaluation studies, the treatment is usually a policy intervention such as an active labour market policy.

Let us define an individual as the unit that will correspond to the unit of randomization in the next sections. It could be a person, a firm, a school or any organization. The researcher is interested in testing the effectiveness of a specific treatment or intervention, T_A, with $T_A = 0$ if the individual is not treated and $T_A = 1$ if the individual is treated. The level y achieved by individual i is a function of the treatment:

$$y_i(T_{Ai}) = \beta_{0i} + \beta_{Ai}T_{Ai} + u_i, \tag{1}$$

where β_{0i} and β_{Ai} are unknown parameters and u_i is an individual specific error term. Obviously there is an issue of selection bias only if there is heterogeneity in the effectiveness of an intervention across individuals (β_{Ai} are individual specific) in a way that cannot be fully controlled for (i.e. there are conditional on unobservable characteristics of the population). In general, β_{Ai} is a function of observable and unobservable characteristics X_i : $\beta_{Ai} = g(X_i, u_i)$. Ignore, for now, the possibility of conducting an RCT. If the intervention was available to everyone, the decision to conduct the intervention would depend on the expected marginal benefit of the intervention. More precisely, suppose individuals have a prior β^*_{Ai} regarding the effectiveness of treatment A. β^*_{Ai} is a draw from a distribution $F(\beta_{Ai})$,and assume that $E(\beta^*_{Ai}) = \beta_{Ai}$ (without loss of generality)[2] and defined on the support $[\beta_{A,\min}, \beta_{A,\max}]$. Suppose conducting the intervention has a fixed cost c_A that we assume identical across individuals.

The expected net benefit of conducting the intervention is $\beta^*_{Ai} - c_A$. Only those with $\beta^*_{Ai} > c_A$ will conduct the intervention. This type of selection corresponds to the "traditional" selection bias usually considered in the policy evaluation literature. Those who would implement the treatment are those with the highest expected marginal benefits of the treatment. Given that priors are correlated with the truth, the estimated average treatment effect $\overline{\widehat{\beta_A}}$ will be a positively biased estimate of $\overline{\beta_A}$.

B. Randomized-Controlled Experiments

The main point of an RCT is to get rid of the selection problem described above. Suppose the researcher is interested in estimating $\overline{\beta}_A$ (the average treatment effect across all individuals). An unbiased estimate of $\overline{\beta}_A$would be obtained by a randomised controlled trial, with a group of N_{Treat} individuals assigned at random to a treatment group (who

[2]Note that the argument carries through even if priors are systematically biased upwards or downwards.

receive treatment A) and a group of $N_{Control}$ individuals assigned at random to a control group (who do not receive the treatment). The key assumption to obtained an unbiased estimate of $\overline{\beta}_A$ is $E(T_{Ai}|u_i) = 0$, that is the assignment to treatment and control is random and is uncorrelated with unobservable characteristics. In that case:

$$\begin{aligned} \overline{\widehat{\beta}}_A &= \frac{\sum_{i \in Treatment} y_i}{N_{Treat}} - \frac{\sum_{i \in Control} y_i}{N_{Control}} \\ &= \overline{y}_{Treat} - \overline{y}_{Control} \end{aligned} \tag{2}$$

C. Outside Option and Competing Interventions

We now discuss the issue of self-selection in a world where possible *competing* and *similar* interventions may exist. For simplicity and without loss of generality, assume there is one possible alternative intervention T_B.

Definition 1. *Intervention T_B is said to be competing with the experimental intervention T_A if it cannot be (practically) implemented in conjunction with the experimental intervention.*

For example, a firm that implements a tournament cannot implement a piece-rate scheme for the same workers.

Definition 2. *Intervention T_B is said to be similar to the intervention T_A if the prior beliefs of the treatment effects of interventions T_A and T_B $(\beta^*_{A_i}, \beta^*_{B_i})$ are positively correlated.*

For example, Belot et al. (2013) consider an intervention consisting of rewarding children for eating fruit and vegetables at lunch. The intervention was conducted at the school level. There are obviously other candidate interventions to increase fruit and vegetable consumption, and schools could hold positively correlated beliefs regarding the effectiveness of these interventions. A special case is if T_A itself is available outside the experimental setting.

In principle, those in the control group are *not treated* and may be involved in other interventions. There are two typical cases. First, participants in the control group are told not to implement T_A or any other treatment that is both competing and similar to T_A. Second, which is typical in medical trials, the control group can continue to take the current best treatment that is available (T_B), but will not have access to T_A. If that is

the case, then the estimated effect will only capture the net difference between the effect of the experimental intervention and the effect of alternative competing interventions. In contrast, non-participants can always implement T_B or even T_A if it is available to them.

D. Participation Decision and Selection Bias

In this section we will derive the participation constraints for the two cases described above. Consider first the situation where $E(\beta_A^*) \geq E(\beta_B^*)$ and $E(\beta_A^*) \geq 0$, that is, the experimental intervention is, a priori, believed to be more effective than the interventions that are currently available, and T_A is believed to have a positive effect.

Presumably there is a direct cost of participating in the experiment, which we denote k (such as providing support for data collection, etc.) but also a potential subsidy s. Also, implementing intervention T_B has a cost c_B (assumed identical for all individuals).

Denote $\bar{g} = \pi(s_{treat} - k - c_A) + (1 - \pi)(s_{control} - k)$.

Case 1 Intervention T_B is not available to the control group.
Then the decision to participate in the experiment must satisfy:

$$\pi\beta_{Ai}^* + \bar{g} \geq \max\{\beta_{Bi}^* - c_B, 0\}, \tag{3}$$

where β_{Bi}^* is i's prior belief about the effectiveness of T_B.

We now derive the participation constraints. We first start with the optimists (such that $\beta_{Bi}^* \geq c_B$). The decision to participate to the experiment satisfies:

$$\pi\beta_{Ai}^* + \bar{g} \geq \beta_{Bi}^* - c_B \qquad \text{if} \quad \beta_{Bi}^* \geq c_B \tag{4}$$

$$\Leftrightarrow$$

$$\beta_{Bi}^* \leq \pi\beta_{Ai}^* + \bar{g} + c_B \qquad \text{if} \quad \beta_{Bi}^* \geq c_B \tag{5}$$

Without loss of generality, let us take the example where β_{Bi}^* is a linear projection of β_{Ai}^*, i.e. $\beta_{Bi}^* = \rho\beta_{Ai}^* + \eta_i$, with ρ being a fixed parameter and η_i an error term, with mean zero and variance σ_η^2. We can then write condition [4] as:

$$(\rho - \pi)\beta_{Ai}^* \leq \bar{g} + c_B - \eta_i \qquad \text{if} \quad \beta_{Bi}^* \geq c_B \tag{6}$$

If $\rho > \pi$:

$$\beta_{Ai}^* \leq \frac{\bar{g} + c_B - \eta_i}{(\rho - \pi)} \qquad \text{if} \quad \beta_{Bi}^* \geq c_B \tag{7}$$

If $\rho > \pi$, then there will be a negative selection bias. All else equal, the higher the correlation between T_A and T_B the more negative selection there is.

If $\rho < \pi$, then positive selection could take place:

$$\beta^*_{Ai} \geq \frac{\bar{g} + c_B - \eta_i}{(\rho - \pi)} \qquad \text{if} \quad \beta^*_{Bi} \geq c_B \tag{8}$$

In contrast, the participation constraint for pessimists (those such that $\beta^*_{Bi} < c_B$) satisfies:

$$\begin{aligned} \pi\beta^*_{Ai} + \bar{g} &\geq 0 \qquad \text{if} \quad \beta^*_{Bi} < c_B && (9)\\ &\Leftrightarrow && (10)\\ \beta^*_{Ai} &\geq -\frac{\bar{g}}{\pi} && (11) \end{aligned}$$

These two conditions show that positive and negative selection can take place at the same time. Without further assumptions on the distribution and joint distribution of β^*_{Ai} and β^*_{Bi}, we cannot draw conclusions on the direction of the bias.

Case 2 Intervention T_B is available to the control group.

This second case corresponds to Malani (2008) implies a lower implicit cost of participating in the experiment. But more importantly, the implications for selection are very different.

Here the decision to participate to the experiment must satisfy:

$$\begin{aligned} \pi\beta^*_{Ai} + (1-\pi)\max\{\beta^*_{Bi} - c_B, 0\} + \bar{g} &\geq \max\{\beta^*_{Bi} - c_B, 0\}, && (12)\\ \pi\beta^*_{Ai} + (1-\pi)(\beta^*_{Bi} - c_B) + \bar{g} &\geq \beta^*_{Bi} - c_B \qquad \text{if} \quad \beta^*_{Bi} - c_B \geq 0 && (13)\\ \pi(\beta^*_{Ai} - \beta^*_{Bi} + c_B) + \bar{g} &\geq 0 \qquad \text{if} \quad \beta^*_{Bi} - c_B \geq 0 && (14)\\ \pi(\beta^*_{Ai} - \beta^*_{Bi}) &\geq -c_B - \bar{g} \qquad \text{if} \quad \beta^*_{Bi} - c_B \geq 0 && (15) \end{aligned}$$

If $\beta^*_{Ai} > \beta^*_{Bi}$, which we have assumed is true in expectations, then those who are optimistic about A will opt in, then the constraint is always satisfied if $c_B + \bar{g} > 0$, which is likely to be the case in a typical randomised controlled experiment. In that case, no negative selection takes place.

However, if the non-participants have access to T_A, while the control group does not (and only has access to T_B), then the participation condition becomes:

$$\begin{aligned}
\pi\beta^*_{Ai} + (1-\pi)(\beta^*_{Bi} - c_B) + \bar{g} &\geq \beta^*_{Ai} - c_A \qquad \text{if} \quad \beta^*_{Ai} - c_A \geq \beta^*_{Bi} - c_B \geq 0 \\
\beta^*_{Ai} &\leq \frac{c_A + \bar{g}}{(1-\pi)} + (\beta^*_{Bi} - c_B) \\
(1-\rho)\beta^*_{Ai} &\leq \frac{c_A + \bar{g}}{(1-\pi)} - c_B + \eta_i,
\end{aligned}$$

such that the most optimistic will select out to implement T_A outside the experiment.

There could still be positive selection in the case where B is not attractive to non-participants or participants in the control group.

$$\begin{aligned}
\pi\beta^*_{Ai} + \bar{g} &\geq 0 \qquad \text{if} \quad \beta^*_{Bi} - c_B < 0, \\
\beta^*_{Ai} &\geq -\frac{\bar{g}}{\pi} \qquad \text{if} \quad \beta^*_{Bi} - c_B < 0.
\end{aligned}$$

3. Conclusion

This paper discusses the implications of self-selection into randomized controlled field experiments. We point out that in many situations; alternatives to the experimental intervention (or the intervention itself) are available outside the experiment. The implication is that being part of the control group entails an opportunity cost, which could lead to both positive and negative selection at the same time.

References

Allcott, H. and S. Mullainathan (2012), External Validity and Partner Selection Bias, *NBER Working Papers* 18373

Belot, M. and J. James (2013), Partner Selection into Policy Relevant Field Experiments, *mimeo*

Gautier, P. and B. van der Klaauw (2012), Selection in a field experiment with voluntary participation, *Journal of Applied Econometrics* 27, 63-84.

Holt, Charles, A. (2005), *Markets, Games and Strategic Behavior: Recipes for Interactive Learning.* Addison-Wesley, Boston

Heckman, James and E.J. Vytlacil (2006), Econometric Evaluation of Social Programs. *Handbook of Econometrics*, Volume 6, James Heckman and E. Leamer, eds., Amsterdam: North Holland.

Heckman, James, R. Lalonde and J. Smith (1999), The Economics and Econometrics of Active Labor Market Programs, *Handbook of Labor Economics*, Volume 3, Ashenfelter, A. and D. Card, eds., Amsterdam: Elsevier Science.

List, John A. & Rasul, I. (2011), Field Experiments in Labor Economics, *Handbook of Labor Economics*, Elsevier .

Malani, A. (2008), Patient enrollment in medical trials: Selection bias in a randomized experiment, *Journal of Econometrics* 144(2), 341-351, June.