Current developments in cluster randomised trials

What can we learn from Economics?

Josselin Thuilliez

Centre National de la Recherche Scientifique (CNRS) Centre d'économie de la Sorbonne-UMR 8174



November 12 2019

Briefly, what am I doing?

- CNRS researcher, economist by training.
- Health and Development Economics mixed with other disciplines: links between Diseases, Income (poverty, well-being), Human Capital. ⇒ Behaviors and Policy implications.
- Examples:
 - Micro Empirical works: disease and Human capital (quasi-experiments, cohort studies, RCTs & cluster RCTs), tests of public interventions (perverse incentives generated by health education and information).
 - Macro Empirical works: social gradient, mortality inequality, economic activity and health (micro-macro works).
 - Theoretical works: economic epidemiology and health traps, relative price of treatment/prevention and disease dynamics, innovations and aid related to disease control.

Objectives today

- Explain how this study design the cluster RCT which is common to you, is used in the same or different way in a context that will be different from a clinical research by people with a different background.
- Focus on concrete examples: different questions/angles, methodological considerations, and comparisons with what I know from your field.
- Provide an overview of specific techniques and critical perceptive from an economist's point of view.

Timely Topic!



Timely Topic!

• Analysis of poverty

- "For their experimental approach to alleviating global poverty."
- But not only, their works cover a **broad range of applications** and research areas:
 - Behavioral economics, information, nudges
 - Health economics
 - Social and political studies: discrimination, votes,...
 - Agriculture, Productivity, Markets
- Some of these areas were using experiments for a long time, but the high number of publications following this "credibility revolution" is unprecedented.
- Led to a large debate on the **use of experiments** in human sciences, as a complement or substitute to other methods, and possible exploitation of results by policy makers.

Nobel Prize 2019

Timely Topic!

BROOKINGS

POLICY 2020 CITIES & REGIONS GLOBAL DEV INTLAFFAIRS U.S. ECONOMY U.S. POLITICS & GOVT MORE



History of field experiments (from an economic point of view): Levitt and List (2009, Eur. Econ Review)

- Birth of field experiment in the 1920s and 30s: Fisher and Neymans works on agricultural experimentation.
- Large-scale social experiments:
 - Early definition: Ferber and Hirsch (1982) "publicly funded study that incorporates a rigorous statistical design and whose experimental aspects are applied over a period of time to one or more segments of a human population, with the aim of evaluating the aggregate economic and social effects of the experimental treatments."
 - **1960s debate over the welfare system**: the New Jersey income maintenance experiment considered to be the first large-scale social experiment conducted in the US (Ross, 1970): positive work incentive?
 - **1980s debate** between experimental advocates and those favoring structural econometrics using naturally-occurring data.
 - **2000s** large-scale examples: PROGRESA (Schultz, 2001), Conditional cash transfer (CCT) and Mexican social policy.

History of field experiments (from an economic point of view): Levitt and List (2009, Eur. Econ Review)

• Current generation of field experiments:

- Carry out this randomization on naturally-occurring populations in naturally-occurring settings (with or without informed consent)
- More ambitious theoretical goals than social experiments (which largely aim to speak to policy makers). Modern field experiments in many cases are designed to:
 - test economic theory,
 - collect facts useful for constructing a theory,
 - and organize data to make measurements of key parameters, assuming a theory is correct.
- Field experiments can help to provide the necessary behavioral principles to permit sharper inference from laboratory or naturally-occurring data.

Different registry system

AEA Registry

AEA RCT Registry
The American Economic Association's registry for randomized controlled trials
Sout Registration Guidelines FAQ Advanced Search

We are happy to announce that all trial registrations will now be issued DOIs (digital object identifiers). For more information, see here.

AEA RCT Registry currently lists 2966 studies with locations in 142 countries.

MOST RECENTLY REGISTERED TRIALS

Reducing distortions in electricity demand

LAST REGISTERED ON OCTOBER 23, 2019

Pay as you go (PAYGo) solar home systems are a market-based technology to increase rural electification for low income households. However, the setting in which consumers use the PAYGo contract may feature market frictions that push consumers away from their optimal demand for electricity. I partner with a solar company in Rivanda to experimentally reduce relevant market frictions. I use this experiment to better understand non-price determinants of demand for electricity among rural, low income consumers.

Celebrity endorsement in promoting pro-environmental behavior

LAST REGISTERED ON OCTOBER 23, 2019

There is greatly growing attention to tacke plastic pollution issues. Previous studies proposed various instruments dealing with the plastic problems. Education and information programs are frequently used and examined as a mean promoting pro-environmental behavior. Several different forms of pro-environmental campaigns have been observed, i.e., public engagement activities with the

AEA Registry

- Amongst the 3004 trials currently registered, typing the keyword "cluster" gives 431 results (\pm 14%).
- Typing the full "cluster randomized trial" provides 222 results (\pm 7%).
- Types of clusters (most often geographic clusters):
 - Firm clusters
 - Workplace
 - Groups of Facebook users
 - Schools
 - Communities, villages, etc...
- In general more than just two arms.
- Typing "stepped wedge" gives only 8 results.

Guidelines

- To my knowledge: no formal guidelines website or checklist as the CONSORT for example.
- Most guidelines are provided in Handbooks or books:
 - The Econometrics of Randomized Experiments. Athey, Imbens. (2017). Handbook of Economic Field Experiments, Volume 1.
 - Using Randomization in Development Economics Research: A Toolkit. Duflo, Kremer, Glennerster (2008). Handbook of Development Economics, Volume 4.

Guidelines

- Organizations also have their own guidelines and evaluation forms:
 - Strategic Impact Evaluation Fund-World Bank
 - Poverty Action Lab (J-PAL) and Innovations for Poverty Action (IPA)
- As a result, practices are often different and guided by the usual structure of economic papers.
- My perception of guidelines in other fields is that existing guidelines for RCT protocol content vary substantially in their recommendations.

Ethics: Informed consent in Social Science (List, Science 2009)

- In most experiments: subjects aware that they are taking part in an experiment and sign consent forms (Nuremberg code).
- However, certain cases in which adhering to rigid ethical rules can affect the very issue that is being studied ⇒ difficult to conduct the research: "if one were interested in exploring whether, and to what extent, race or gender influences the prices that buyers pay for used cars, it would be difficult to measure accurately the degree of discrimination among used car dealers who know that they are taking part in an experiment."
- "Ethical issues surrounding human experimentation are of utmost importance. Yet, the benefits and costs of informed consent should be carefully considered in each situation. Those cases in which there are minimal benefits of informed consent but large costs are prime candidates for relaxation of informed consent."

Philosophical considerations?

A different perspective?

- Of course questions are different in economics compared to biostatistics, epidemiology or clinical trials.
- The historical development of the discipline is also different:
 - Initially, the limitations of nonexperimental methods were the topic of the rich and active field of econometrics.
 - Most indices of causality are similar (strength, consistency, specificity, temporality, gradients) but plausibility is not biological. Consistency may also be more variable and context-dependant (problem of external validity.
 - However, as for biological models, without theoretical models, it is often hard to have ex-ante assumptions on plausibility or extract testable rules on causality/multicausality.
 - Behavioral lab experiments may help but it is not granted that people will behave similarly
 - Counter-intuitive results are very frequent.

Content

- I will give concrete examples of RCTs so that you can have an idea of what we do in our field. Inside them I will give one personal example.
- I will next discuss some techniques and developments that are probably more used in economics because of the historical development of the discipline.
- RCTs and cluster RCTs receive regular critical attention. I will expose some of the critics that are specific to economics.

The use of RCTs and cluster RCTs in empirical economics: a brief overview from recent seminal papers

A few seminal papers

- Banerjee, Duflo, Glennerster, and Kothari. (BMJ, 2010). Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives.
- **Duflo & Saez (2007, Quarterly J. of. Economics)**, The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment.
- Kremer & Miguel (Econometrica, 2004), Worms: Identifying impacts on education and health in the presence of treatment externalities.

- Published in **BMJ**, very close to clinical trials standards. 9 pages, BMJ format.
- **Objective**: assess the efficacy of modest non-financial incentives on immunisation rates in children aged 1-3 and to compare it with the effect of only improving the reliability of the supply of services.
- Design/Methods: Clustered randomised controlled study.
 - 134 villages randomly selected from a Seva Mandir catchment area in Udaipur (India) with 30 HHs.
 - 3 arms (A/ once monthly reliable immunisation camp (30) B/ once monthly reliable immunisation camp + non monetary incentives (30) C/ Control (74)).
 - **Outcome**: Proportion of children aged 1-3 at the end point who were partially or fully immunised.

- Ethical approval: approved by the health ministry of the government of Rajasthan, the office on the use of human subjects at Massachusetts Institute of Technology, and the ethics committee of Vidhya Bhawan, the university which hosted the project in Udaipur.
- **Informed consent**: was first obtained orally at the community level from the research villages through village meetings to which all adult members of the village were invited. Individual level informed consent was then obtained orally from every family participating in the study.

- **Power calculation**: Taking into account correlation of the end point within a village and clustering of the treatment at that level (a intracluster correlation of 0.25 was assumed based on a preliminary survey) and given a baseline immunisation rate of 2% in the control group, we determined that a sample of 30 villages per treatment arm, with a random sample of 30 households per village (assuming about 1.4 children aged 1-3 years surveyed in each household), was sufficient to obtain 80% power for a 5% level test of a difference of at least five percentage points in the probability of being fully immunized between any two groups (treatment A, treatment B, and comparison). **The** larger control group increases power.
- Baseline characteristics are provided in the paper.

• Analysis:

- ITT analysis, no control variables.
- **Count variables** (number of immunisation): report values in the treatment group, difference across groups, and relative risks. The analysis adjusts for clustering at the village and the family level.
- **Difference in proportion**: a multilevel mixed effect linear model of the probability of being immunized on the treatment indicator, with a hierarchical error structure that allows cluster level heterogeneity (random effect) at the village and at the family level.
- **Relative risk**: multilevel mixed effect Poisson model with the same hierarchical error structure.

• Presentation of results

- Primary end point: impact on immunisation in treatment village.
- Probably one difference: impact on neighbouring villages is provided (externalities! villages within a few kilometres).
- A short cost analysis is provided for CEA.
- Summary, Limitations of the study, policy implications are given..



• Main Conclusions

- In a setting with a low immunisation rate (under 6%), improving the reliability of services modestly improved uptake of immunisation.
- Small non-financial incentives, combined with improved reliability, had large positive impacts on the uptake of immunisation and were more cost effective.

Duflo and Saez (2007), Quarterly J. of. Economics

- Format is different. 29 pages in QJE format (top 5 journal in economics).
- **Objective**: shed light on the role of information and social interactions in employees decisions to enroll in a Tax Deferred Account (TDA) retirement plan within a large university. Peer effects.
- **Randomization in two steps**: departments (*clusters*) within a university were selected (following a stratified procedure) and next randomized to receive treatment. Next inside treated departments, part of the staff received a letter.
 - Employees are potentially subjected to two kinds of treatments: they can receive the invitation letter themselves, or they can be in a department where some employees received the letter.
 - In the presence of social interactions, employees who work in departments where some people received the letter can be affected by the experiment even if they did not receive the letter themselves.

Large scale RCTs

Duflo and Saez (2007), Quarterly J. of. Economics

- The experiment: encouraged a random sample of employees in a subset of departments to attend a benefi?ts information fair organized by the university.
- **Outcome**: Attending the fair and TDA enrollment.
- No ethical approval is mentioned in the article.
- Analysis: Reduced form estimates through OLS. Regressions control for department, gender, year of service, age, and salary. Standard errors are corrected standard errors for clustering at the department level
 - Average effect of department treatment
 - Effect of letter and department treatment
 - Next the authors complexify the framework to account for different mechanisms: differential treatment effects, social network effects, and motivational reward effects.

Duflo and Saez (2007), Quarterly J. of. Economics

• Findings:

- The experiment multiplied by more than five the **attendance rate** of treated individuals (relative to controls), and tripled that of untreated individuals within departments where some individuals were treated.
- **TDA enrollment** five and eleven months after the fair was significantly higher in departments where some individuals were treated than in departments where nobody was treated.
- However, the effect on TDA enrollment is almost as large for individuals in treated departments who did not receive the encouragement as for those who did.

Large scale RCTs

Kremer and Miguel (2004), Econometrica

- Format is different. 58 pages in Econometrica format (top 5 journal in economics)! Much longer.
- **Objective**: evaluate a Kenyan project in which school-based mass treatment with deworming drugs was randomly phased into schools, rather than to individuals, allowing estimation of overall program effects. Take externalities into account.
- Design/Methods: original research design was based on a stepped-wedge analysis.
 - 3 arms (groups) of 25 schools/arm (A- free deworming treatment in both 1998 and 1999; B- 1999 only; C-Treated in 2001). About 400 pupils per group. 80% of children in treatment groups actually received treatment.
 - "In 1998, Group 1 schools were treatment schools, while Group 2 and Group 3 schools were comparison schools, and in 1999, Group 1 and Group 2 schools were treatment schools and Group 3 schools were comparison schools."

- Analysis presented as "Econometric Specifications" or "estimation strategy"
 - Standard ITT is used but in addition to this they estimate cross-school externalities by taking advantage of variation in the local density of treatment schools induced by randomization.
 - Estimate program impacts in treatment schools, as well as cross-school treatment externalities.

(1)
$$Y_{ijt} = a + \beta_1 \cdot T_{1it} + \beta_2 \cdot T_{2it} + X'_{ijt} \delta + \sum_d (\gamma_d \cdot N^T_{dit}) + \sum_d (\phi_d \cdot N_{dit}) + u_i + e_{ijt}.$$

where T_{1it} and T_{2it} are indicator variables for school assignment to the first and second year of deworming treatment, respectively; and X_{ijt} are school and pupil characteristics. N_{dit} is the total number of pupils in primary schools at distance d from school i in year t, and N_{dit}^{T} is the number of these pupils in schools randomly assigned to deworming treatment.

Josselin Thuilliez (CNRS)

• Health cost-effectiveness:

- Taking into account externalities (untreated on treated and treatment on control across and within schools) the total number of DALYs averted as a result of the program is 649, which translates into a cost of approximately USD 5 per DALY saved.
- Externalities account for 76% of the DALY reduction.
- And 99% of the total DALY reduction is due to averted schistosomiasis, a disease caused by parasitic worms of the Schistosoma type (not geohelminthes).
- Consequently, if externalities are not taken into account or if disease specificity is not taken into account, the program would not meet strict cost-effectiveness criteria in the poorest countries.

• Education cost-effectiveness:

- Deworming increased school participation by 0.14 years per treated child or USD 3.50 per additional year of school participation, including both effects on the treated and externality benefits..
- Comparatively educational programs cost USD 99 per additional year of participation: *evaluated textbook provision, grants to school committees, training for teachers, and incentives for teachers based on student test scores and dropout rates, paying for the uniforms.*
- Potentially huge Labor market benefits and wages effect: deworming increases the net present value of wages by over USD 30 per treated child at a cost of only USD 0.49.
- As a result, a large government deworming subsidy is optimal, considering all the externalities and the CEA.

32 / 61

Worms War: Kremer and Miguel (2004), Econometrica



• Replications results:

- Hicks, Joan Hamory, Michael Kremer, and Edward Miguel. 2015. Commentary: Deworming externalities and schooling impacts in Kenya: a comment on Aiken et al. (2015) and Davey et al. (2015). Int. J. Epidemiol. (2015): 1-4.
 - Only small changes in estimated coefficients. Aiken et al. (2014) also find externality effects within schools on untreated pupils, as well as externality effects across schools up to 3 km away.
 - The debate is on schools located from 3 to 6 km. No evidence that worm infection externalities extend beyond the 12 closest schools to the full set of schools within 6 km.
 - A peculiar weighting procedure and assumptions behind it is debated.
- Data and all programs are fully available for replication.
- 10 years after: "Worms at Work: Long-Run Impacts of a Child Health Investment" QJE 2016.

Of course, many more papers

- Banerjee et al. (2017, J. of Devt Econ.). Can Iron-Fortified Salt Control Anemia? Evidence from Two Experiments in Rural Bihar.
- Kremer and Miguel (2007, QJE). The illusion of sustainability.
- Crépon, Dulfo et al. (2013, QJE). Do Labor Maket policies have displacement effects? Evidence from a clustered randomized experiment.
- Das et al. (including Banerjee) (2016, Science). The impact of training informal health care providers in India: A randomized controlled trial.

• • •

My own modest experience: an example from a cluster RCT in La Réunion

• Thuilliez & Dumont (World Bank Economic Review, 2019) Public Mosquito Abatement: A Cluster Randomized Experiment.

• Objective:

- Build a theoretical model of mosquito abundance and public and private abatement response.
- Test the model with a population based RCT: testing the effect of public intervention on private protective behaviors.
- The results show that public intervention may paradoxically increase risk which is puzzling!
- Methods: Model + Cluster RCT.

- Simple model underlying the experiment (2 households, 1 public agency).
 - Number of mosquitoes, Mi, =decreasing and convex function of public intervention, Pub, and private elimination of breeding sites by 2 households, *Private_i* and *Private_{-i}*. By including *Private_{-i}*, we incorporate an **externality**.
 - Households, which are endowed with income, Y, face positive prices of public intervention (pPub) through taxes, individual behavior (pPriv) that have an opportunity cost or a direct cost, and consumption of insecticides (pc).
 - Thus, each household i wants to minimize the mosquito nuisance and pc, which is equivalent to solve the following utility minimization problem:

$$\min_{\mathsf{Priv}_i,c_i} \left(M_i(\mathsf{Priv}_i,\mathsf{Priv}_i,\mathsf{Pub})) + c_i \right), \tag{1}$$

• Lagrangian and Kuhn-Tucker theorem.

- We derive some comparative statics that motivate the randomized controlled experiment. We treat Household 2 as the rest of the zone.
- Several situations may be derived from this simple model
 - Case 1: if public and individual private mechanical elimination are substitutes
 - **Case 2:** If zone-level public and individual private mechanical elimination are complements
 - And for each case, if the elimination levels among neighbors are **a**) complements or **b**) substitutes

- We derive some comparative statics that motivate the randomized controlled experiment. We treat Household 2 as the rest of the zone.
- Cases 1(a), 1(b) and 2(a) are consistent with public abatement actually making mosquitoes more abundant.
- In the presence of an externality and strategic complementarities, aggregate relationships will overstate individual elasticities and exaggerate the private response to public abatement.

- Empirical test of the model with a cluster RCT
 - 26 zones randomized: 13 treated, 13 control.
 - Intervention was an intervention of the Agence Régionale de la Santé: elimination of larval gites, health education.
 - Balance at baseline in a broad range of statistics between the treated and control groups (we had 2 surveys).
- Outcomes enable to check for both mosquitoes abundance and private behaviors and other mechanisms.
- Several types of analyses performed through the publication process: ITT through simple OLS, mixed models, or 2SLS for LATE.

	(1) Overall Standardized index	(2) Tend the garden	(3) Clean green waste in the garden (fruit, branches, etc)	(4) Clean waste dumped in the garden	(5) Cover or clean water reservoirs	(6) Clean water reservoirs daily or weekly
ITT	-0.107*	-0.051	-0.022	-0.072**	-0.049**	-0.037
	(0.055)	(0.031)	(0.030)	(0.030)	(0.021)	(0.036)
R2	0.03	0.026	0.034	0.033	0.024	0.026
2SLS	-0.128*	-0.061	-0.026	-0.086**	-0.059**	-0.044
	(0.067)	(0.038)	(0.036)	(0.039)	(0.025)	(0.042)
Observations	1043	1043	1043	1043	1043	1043

	Stegomya Indices				
	(1) Infested Household	(2) Proportion of infested containers per Household	(3) Number of infested containers per Household		
ITT	0.087**	6.148**	0.259**		
	(0.039)	(2.578)	(0.103)		
R2	0.083	0.071	0.05		
2SLS	0.105**	7.372**	0.311**		
	(0.047)	(3.109)	(0.126)		
Observations	1043	1043	1043		

Motivations and specific techniques in econometrics

Athey and Imbens (2017)

Motivations for cluster RCTs in economics

- Given a fixed sample size, this design is in general not as efficient as a completely randomized design or a stratified randomized design.
- The motivation will be different:
 - Interference between units at the unit-level (no account for the within-cluster interactions).
 - Easier to sample units at the cluster level.
- Recommendations for analysis:
 - Cluster-level analysis is recommended as more transparent and more directly linked to the randomization framework.
 - However, unit-level analysis allows the analyst to impose additional modeling assumptions (e.g. restricting the effect of covariates to be common across clusters).
 - In practice as you saw, several options are used. While mixed/hierarchical models are used for journals such as BMJ, Lancet, Science, they are much less used in Economic Journals. I think the reason for that is simply historical practice.

IVs, LATE and Partial Compliance



• Instrumental variables models (pseudo-randomization):

- An instrument is a variable that only affects the outcome via the treatment variable, not directly.
- Hence, if the outcome moves when the instrument moves, it must be due to the treatment.
- By dividing the variation in outcomes induced by the instrument by the variation in treatment induced by the instrument , we can measure the effect of the treatment.

Josselin Thuilliez (CNRS)

IVs, LATE and Partial Compliance

- In a randomized trial, the assumptions are met by design in the act of randomization.
- In cases of treatment-effect homogeneity, IV studies estimate the effect on the marginal subject, the average treatment effect for patients whose treatment was determined by the instrument.
- Techniques used include 2SLS, LIML, GMM, ...
- See for instance Imbens, G. W., and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects (LATE)", Econometrica. Or Complier average causal effect (CACE).
- Sussman et al. (2010, BMJ): An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials.

Externalities: A central concept

- In economics, externalities are unexpected and indirect effects of consumption or production activity that cannot be quantified through prices on a usual market.
- Examples of externalities include the social benefit of vaccination, use of pesticides, education and friends,...
- With respect to social externalities, Manski (1993, RESTUD) is an important paper in this area: individuals are exposed to similar environmental stimula as their peers. The interactions may be a nuisance that affects the ability to do inference.
- Active field of research in economics.

Testing for Treatment Effect Heterogeneity

- In many cases researchers are interested in the presence or absence of heterogeneity in treatment effects.
- For example, one may be interested separately in the effect of Northern versus Southern areas.
 - Analyze separately the two samples is one option.
 - Evaluate heterogeneity with respect to each covariate, one by one.
 - Multiple testing problem: List, Shaikh, and Xu (2016) propose a computationally feasible approach by bootstrapping, and accounting for correlation among test statistics.
 - Other methods include lasso-like methods, machine learning, "regression tree" or "recursive partitioning" methods (Athey and Imbens, 2016).
- This may also be useful to provide recommendations for other settings.

Randomized saturation

- Under such designs, the proportion of people assigned to treatment in a given area is randomized, as well as the treatment status of individuals within such area (as in Duflo and Saez, 2007).
- But the framework can include more groups with different probabibilities of treatment within the treated clusters.
 - Malani (2006, J. of . Pol. Econ) Identifying placebo effects with data from clinical trials: DBRCT (Double blind RCT) fail to fully account for the efficacy of treatment if there are interactions between treatment and behavior, for example, if a treatment is more effective when patients change their exercise or diet.
 - An example of solutions includes 2 by 2 trials



Source: Chassang et al. Plos One 2015.

Program Evaluation in the Presence of Strategic Interactions (Acemoglu et al., WP 2017)

- Recent improvements in program evaluation techniques have allowed researchers to estimate the spillover effects of programs and policies, in addition to the direct effects.
- In some settings, there may be an interaction between the direct effects and the spillover effects of a treatment if the size of spillovers depends on an individual's own treatment status.
- These interactions are strategic if an individual's treatment status depends on the treatment of their neighbors.
- In the presence of strategic interactions, reduced form estimates of direct effects are biased, even when a 'randomized saturation' experimental design is used.

Using RCTs to Estimate Long-Run Impacts in Development Economics

- Novel opportunity to generate high-quality evidence on the long-run drivers of living standards and economic productivity (in poor countries).
- Limited number that have been followed up to date.
- Improved participant tracking methods; alternative research designs; and access to administrative, remote sensing, and cell phone data.

Using RCTs to Estimate Long-Run Impacts in Development Economics



Figure 1

Cumulative number of completed randomized controlled trials (RCTs) in low- and middle-income countries from 1995 to 2015 in the American Economic Association's RCT Registry (https://www. socialscienceregistry.org).

Critical Perspective

Heckman (Nobel Prize 2000)

• "The Scientific Model of Causality"

- Counterfactual statements have to be made within a precisely stated model.
- Ambiguity in model specification implies ambiguity in the definition of counterfactuals and hence of the notion of causality.
- Criticisms of Social Experiments:
 - Experiments Provide Little Evidence on Many Questions of Interest
 - The Intrinsic Variability in Evidence from Randomized Experiments
 - Randomization Bias
 - Institutional Limitations
 - Substitution Bias



Deaton (Nobel Prize 2015)



- "All that said and done, RCTs are still generally the best that can be done in estimating average treatment effectsand in warranting causal conclusions."
 ⇒ "It is this claim that is the monster that seemingly can never be killed" (Deaton and Carwight, 2018)
- "Experiments have no special place in the hierarchy of scientific evidence seems to me to be clear" (Sampson, 2018).

Internal and external validity

- Internal validity: refers to the ability of a study to estimate causal effects within the study population. Outliers? Heterogeneity? Asymmetric distribution of effects?
- External validity: generalizing causal inferences, drawn for a particular population and setting, to others.
 - Can refer to the population under study. Is it replicable to different places?
 - Multicentric studies are made in economics as well: Banerjee et al. (MIT WP, 2015) A multifaceted program causes lasting progress for the very poor: Evidence from six countries.
 - External validity can also refer to the type of medical treatment studied. For instance, is a subsidy that is effective for bednet to prevent malaria also effective for other preventive measures ?

External validity



Source: Bates et al. Field Action Science Report.

Conclusions

What can be learned from economics?

- The spread and use of cluster RCTs being much more recent in (development) economics than in psychology or medicine, economics still has much to learn from these disciplines.
- However the particular nature of social interventions may give a specific attention to methodological research that are not explored elsewhere.
 - No placebo or difficult to say if placebo effect is due to the placebo or externalities.
 - Difficult to render the experiments fully double blind.
 - Moral issues discussed by Deaton on poverty (very often rich scientists analyzing poor people, the reverse being impossible).

Next steps...

- Trying to mix clinical trials (randomizing medical treatment or preventive measures) and social/field experiment will become frequent.
- Initially, when working with epidemiologists or clinicians, the added value of our approach was not directly perceived.
- Wrong belief that cost-effectiveness calculations, or "socio-economic" surveys is our specialty instead of behaviors and mechanisms.
- Interdisciplinary culture is difficult to build and sustain.
- This conference is a good evidence of change!

Thank you for your attention!