WHY IS GLOBAL POVERTY SO HARD TO MEASURE? AND SO HARD TO REDUCE?

Angus Deaton, Princeton
Iseo, June 15th, 2016

Global poverty

- The World Bank: "Our dream is a world free of poverty"
- MDG1: Eradicate Extreme Poverty and Hunger
  - Halve, between 1990 and and 2015, the proportion of people whose income is less than $1 (1993 PPP) a day
- SDG1: End poverty in all its forms, everywhere
  - By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day
  - World Bank interprets this as less than 3 percent by 2030
- President Obama in SOTU 2013
  - "The United States will join with our allies to eradicate such extreme poverty in the next two decades."
  - New goal for US, and USAID in particular, which had not previously targeted global poverty

Why does global poverty exist?

- More specifically, why have World Bank and other MDBs and NGOs, not eliminated poverty already?
- Task is perhaps not such a large one
- Calculation from Great Escape, see next slide
- Some, e.g. Thomas Pogge, that "we" really don't want to, or care
- Others, that aid (or altruism) has not been effective
  - Evidence based policy has been absent: too many fads, or politiics
  - Propose greater use of randomized controlled trials
    - "Britain has given the world Shakespeare, Newtonian physics, the theory of evolution, parliamentary democracy. ….. and the randomized trial" (*BMJ*)
  - "The World Bank is finally embracing science" Lancet editorial, 2004
  - But WB gave up specific projects because they didn't work if macroeconomic conditions were a mess
    - That problem has not gone away, though more politics than macro

Global poverty: magnitudes

- Update a calculation from The Great Escape, Chapter 7
- According to the World Bank, there were 900 million people living below $1.90 a day in 2012
  - From the poverty gap measure, we can calculate that each is $0.56 cents short of the line
  - Need $502 million per day to "eliminate" poverty
  - 503 million people in European Union, 426 million adults
  - 319 million in the US, 240 million adults
  - $0.75 per adult per day (less if we include Japan, Australia, etc.)
- But, $0.75 USD buys about 2.5 times as much in poor countries, by PPP correction
- We need 0.30 cents a day in *actual* USD to eliminate global poverty
  - Cash transfers?
- Actual DAC aid in 2014 was $134.4 billion=0.41cents per poor person per day, or $1.03 in $ PPP in poor countries.

Measurement

Measurement

- We won't know about SDG1 unless we know how many poor people there are
  - Can monitor changes over time
  - Note that SDG1 makes greater demands on measurement than did MDG1
    - If you change the measure, you can still halve the new measure
    - Doesn't work if you are committed to zero, or three percent
- We also can't monitor success of *any* poverty strategy unless we can count
  - This goes beyond the MDGs/SDGs
  - Success of projects, by themselves, is not enough
    - Unintended consequences, GE effects, politics, macro conditions, can mean that projects succeed without reducing poverty

Elements of measurement

- Household surveys where we can count the number of people below the local equivalent of the global line
- A global line, usually in USD, but could be something else, world rupees v international $
  - Rhetorical value to UN and others of USD: success of $1-a-day concepts
- A set of converters so local equivalents of global line have the same purchasing power everywhere
  - PPPs from International Comparison Project

Each element is problematic

- I will discuss each briefly

- Global poverty counts are hypersensitive to difficult-to-resolve measurement issues
- Underlying question:
  - Can we do this well enough to make it worth doing at all?
- Do "we" care?
  - Comparison with domestic poverty measures
  - Political economy is different in domestic v international
  - Who is accountable for success of goals?
    - If no one, then measurement not likely to be well-done
    - Measurement depends on political accountability

Global poverty over time

	Time series of poverty			
	1981	1993	2005	2012
Poverty line PPP date	$1.90 2011	$1.90 2011	$1.90 2011	$1.90 2011
HCR % Millions	53.5 1,982	34.8 1,925	24.6 1,358	14.9 897
East Asia South Asia SS Africa LAC	80.6 58.1 .. 19.7	52.6 47.9 61.1 14.4	18.6 35.0 50.5 9.9	7.2 18.8 42.7 5.6

Effects for 1993 of changing PPPs

		Changing ICPs, all for year 1993			
		1993	1993	1993	1993
Poverty line		$1.01	$1.08	$1.25	$1.90
PPP date		1985	1993	2005	2011
HCR %		29.4	28.2	39.2	34.8
Millions		1,350	1,304	1,799	1,925
East Asia		26.0	25.2	50.8	52.6
South Asia		43.1	42.4	46.9	47.9
SS Africa		39.1	49.7	56.9	61.1
LAC		23.5	15.3	10.1	14.4

Global poverty and PPPs

- Estimates are VERY sensitive to the round of the ICP that we use (as well as to changes in global line that they induce)
- This is true, not just for the total, but for regional breakdown
  - Africa for ICP1993 versus ICP1985
  - South Asia for ICP2005 versus ICP1993
- Why does ICP keep changing these numbers?
  - Invite me back next year
  - Seriously hard problem: quality v representativeness, especially across very different countries
  - ICP is not continuous, and no brake on methodological revisions
- Perhaps
  - take numbers now, and update them using local CPIs, so no use of future ICPs

NAS versus surveys

- Serious conflict between NAS and survey consumption figures, many countries, here is example from India
- In 1972-3, survey mean 5 percent short of NAS mean
- In 1983-4, survey mean 25 percent short of NAS mean
- In 2009-10, survey mean 46 percent short of NAS mean
- With adjustment for differences in concept, 33 percent short
  - "there are infirmities in both sets of estimates"
- India is only one of many countries with this divergence
- Difference in concept is itself an issue for poverty
  - State healthcare provision v private provision, different in different countries
  - What about rents? In NAS not in surveys

Scaling up and the density effect

- Once upon a time in India, survey means were scaled up to match NAS data
  - Some still call for this in India, and in global poverty
  - This makes poverty reduction much more rapid around the world
- Indian poverty in 2012 is 231 million (out of 897 million worldwide)
  - If we were to scale up mean by 25 percent to match NAS, Indian poverty falls to 73 million
    - World's greatest poverty reduction program!!
  - Even a 10 percent plus/minus 305 to 160 million
  - 10 percent is small relative to PPP uncertainty, and survey uncertainty
- These uncertainties almost certainly worse in Africa, though there are fewer people near the global poverty line
- This is the "density effect." Many people near the global line, so small changes have BIG effects
  - 100 million people in [1.8,1.9], 100 million in [1.9, 2.0]

Reporting periods

- ☐ Response to NAS v survey puzzle
- ☐ A good example of how politics penetrates deeply into measurement: the right for NAS, the left for surveys
- ☐ How much rice did you buy over the last XX days?
  - ☐ Indian tradition, from Mahalanobis, was XX=30, based on an experiment
  - ☐ In the debate, right argued this was too long, other statistical offices use 7 days, and that people would forget over 30
  - ☐ This might explain some of the gap between NAS and NSS
    - ■ Though not clear that it can say anything about why that gap increases over time, which is the key issue
- ☐ One side wanted 7 days, other 30 days, unresolvable
- ☐ How to decide? Do a randomized controlled trial, of course!
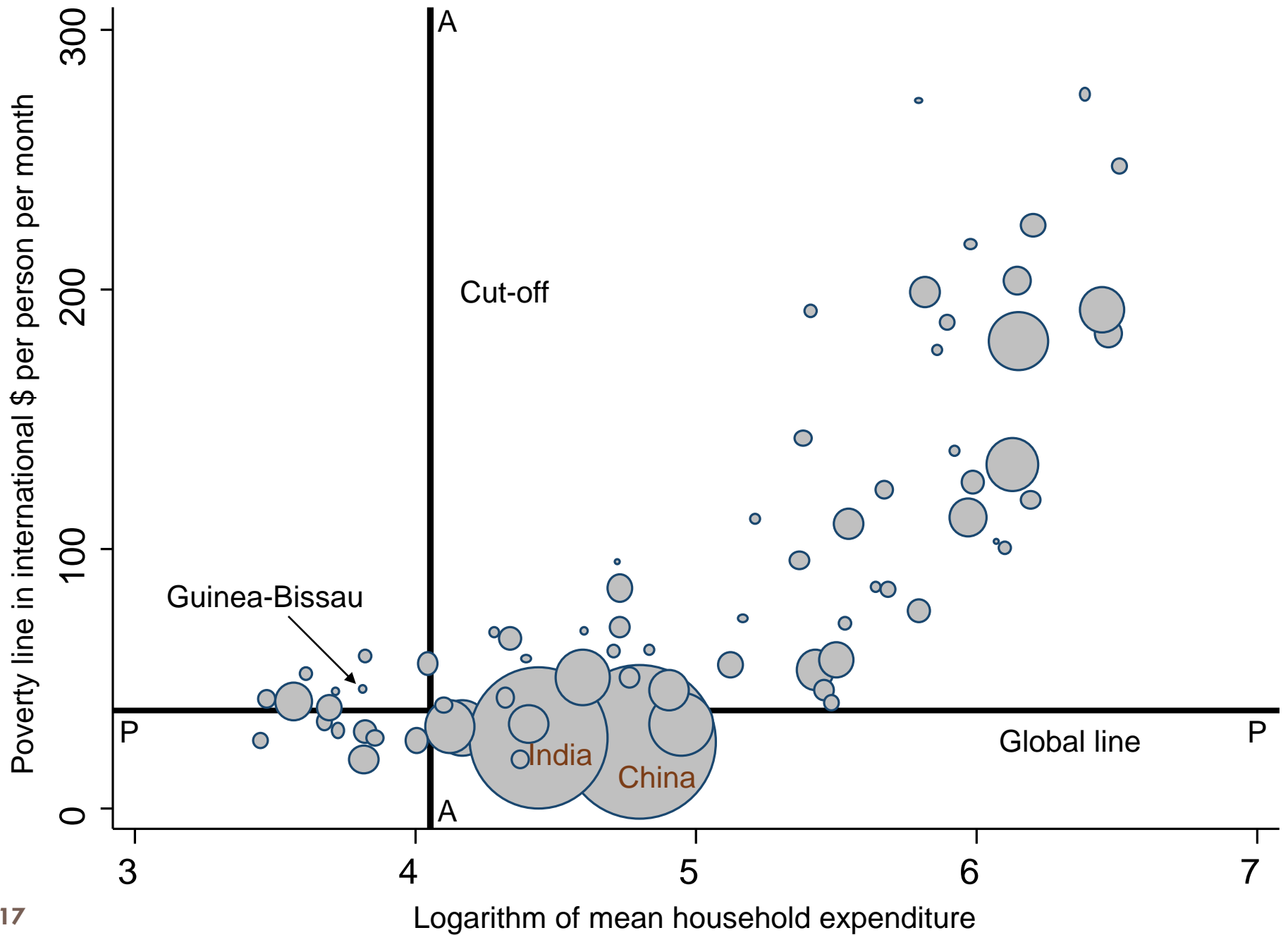
The reporting period debacle

- Surveyors did a nationwide RCT, randomizing 7 and 30 over villages
- RCT showed, indeed, that flow of reported consumption at 7 days was higher than flow at 30 days
  - Triumph for the right, or was it?
- The effect is HUGE: 175 million Indians were removed from poverty in 1998, close to a half
  - This was more than ten percent of WORLD poverty
- But this didn't resolve anything: as always RCTs do not tell us WHY the difference occurred, or which was correct
  - But they did raise the stakes in the debate, because there was a lot to fight over
- Decided to put *both* measures on survey, contaminating both!

The global line

- Is chosen as an average of the poverty lines of the world's poorest countries
  - Done after conversion to PPPs, so depend on PPPs again
- Unweighted average: if they are weighted, then India and China tend to dominate
  - But unweighted means that a change in the line in a tiny country can put hundreds of millions of Indians into or out of poverty
- India became too rich for its poverty line to count, so its line dropped out, which raised the global line, which put millions of Indians into poverty
  - India became poorer because it got richer
- Lots of these lines have very dubious provenance
  - Even when we know what it is

Poverty line in international $ per person per month

Logarithm of mean household expenditure

A — Cut-off
A

P — P
Global line

Guinea-Bissau

India   China

17

Basic needs instead?

- Count the number of people below some calorie cut-off?
  - Poverty lines often originate in such calculations
- But in India and possibly also in China, per capita calorie consumption has been falling for many years
  - Perhaps turned around in the most recent surveys
- So calorie-based counts show Indian poverty *increasing*
  - Unlikely given many other measures
  - Likely because, as people get richer, they do less heavy manual labor, which is an inferior good, or bad
  - So they need fewer calories
  - But we have no direct measures of this

What to do?

- Realistic standard error for global poverty might be 200 million in either direction
- Presumably some autocorrelation in errors, so changes are likely to be a good deal more accurate
- Is this good enough?
- Not all of the problems discussed above are easily fixed

- Use other measures: anthropometrics, infant and child mortality, comparable if imperfect measures, like durable goods?
  - Note that these are NOT the same thing
  - IMR and CMR can fall even when incomes are falling, so they are not proxies but different things
- I think we should put less emphasis on global poverty measures

Finding out "what works"

"If we know, then poverty will be history"

Effective altruism based on this

Are Randomized Controlled Trials a solution?

Randomized controlled trials

- Argued that they give strong evidence of causality, unbiased estimates of the effects of policy, and are largely exempt from many of the problems of standard econometric techniques
- Certainly, they have many strengths, though also many weaknesses
- I will talk about some of these
- I also want to talk about how to USE the results of a good RCT
  - Often, but unhelpfully, called external validity
  - Looking for external validity in the usual sense is hopeless, but short-changes RCTs
- Start with a few observations about problems of doing RCTs

About RCTs

- What they do, and what they do not do
- Do they give the right answer in the setting in which they are run?
  - Internal validity, but better to ask WHO they have been shown to work for
  - In a heterogeneous world, internal validity is not very useful
- I am NOT arguing that RCTs don't work
  - Only that they have their own strengths and weaknesses
  - As do other methods of knowing, including standard engineering and economic expertise, e.g. from the World Bank
  - And they are limited in ways that are important

Building evidence

- The WB used to think it knew the effects of a project
  - But worried about valuing it: needed the right prices
  - Large literature in deriving "shadow" prices
- Now we worry about whether we know the effects of project
  - So we want to test: need quantities
  - "What works" is both quantities and values
  - Still need to value things, sometimes not obvious how to
- RCTs are seen as the best way to gather this evidence
  - Superior to other forms of evidence based on economic or engineering or public health expertise (expertise is challenged)
  - Need no expert knowledge for evaluation by RCTs: anyone can understand
  - This is only an advantage if we think economists don't know anything useful
- Given this, we can eliminate poverty, one project at a time

Getting RCTs right

- Is VERY hard, as anyone who has run one will tell you
- In ideal circumstances, they give an unbiased estimate of ATE
    - Meaning if you did the trial over and over, you would be right on average
    - Why this is relevant is a considerable puzzle
- No presumption that estimate is close to truth
- Indeed randomization artificially *introduces* noise
    - Banerjee, Chassang, and Snowberg theorem: loss minimizing design does NOT randomize
- People think randomization controls for unobservables
    - Which is a misunderstanding
    - But it seems to be a powerful argument for the lay public
    - NOT true that, by randomization, ONLY thing that is different is the treatment
- Without blinding, effect could be ANYTHING
    - Many possible channels of effect other than the effect of the treatment itself

Statistical inference

- Fisher argued for randomization, not because it was precise, but because it allowed him to calculate standard errors
- New and old debates about this
- Difference in two means is a poorly behaved object
- Student t-distribution doesn't apply
- Particularly bad when treatment effects are assymetric: e.g microfinance, Deaton and Cartwright (2016)
- Almost half of all experimental papers in AEA journals find significance that it not really there, Young (2016)
  - Clustering and heteroskedastic consistent estimation can overstate t-values
- As of now, we DON'T have a method of assessing significance in RCTs

Simplicity and credibility

- Many believed that RCTs would cut through complexity of statistical and econometric evidence
- Economists were obsessed with identification and causality
    - Every result was challenged and hard to defend
    - RCTs seemed to offer a solution
- Yet controversy raged on:
    - Worm wars: dozens of papers
    - No agreement even on internal validity
    - Cochrane says no evidence for deworming on education: Givewell disagrees
    - No health effects in million person trial in India, or arguably in original paper
    - Without health as a mediator, how can this possibly work?
    - See Macartan Humphreys' excellent blog

What about replication?

- Surely if it works over and over, we are OK?
  - We have discovered something like the law of gravity
  - But gravity is a much higher-level concept than the experiments we look at
- Bertrand Russell's chicken is fed every morning by the farmer
- Based on repeated evidence, chicken concludes that when she hears the farmer coming, she will be fed
- Inference is great until Christmas day
  - Farmer comes, wrings her neck and eats her
- Chicken did not use an RCT, but we could have done so for her, and inference would be the same
  - Her false inference is not through lack of rigor
- Chicken did not understand economic and social structure underlying the observed regularity
  - "more refined views of the uniformity of nature would have been useful to the chicken" Russell (1912).
- Let's not think like chickens: we have to do serious transportation of results
- External validity cannot be about simple generalization
- Replication is of limited value if we do not understand "the uniformity of nature"

A simple but useful example

- A new fertilizer increases output for treatment cocoa farmers over controls in an RCT, treatment farmers get rich
  - Scale up, price goes down, farmers get poor
  - Opposite sign: causal effect in the opposite direction
- This should *NOT* be seen as a failure of RCT, or a failure of "external validity"
  - The experiment is just fine, and perfectly correct
  - An opportunity to *use* RCT in a broader context
  - This will require observational work and modeling
  - This is not a *disadvantage*: just what it takes to do serious work!
- Going to scale almost always requires this sort of process
  - RCT results need to be set in the context of other things that we know. We have to use them as economists!

Cash transfers

- Imagine a man who moves in next door, with a wife who is impoverished and abused by her husband
  - You feel an obligation to help her
  - Do you give him money? Probably not
  - Do you give her money? Who is likely to get it?
  - Maybe give to him with conditions? Perhaps, but only if you are able and prepared to enforce the conditions
- Imagine a better neighbor, who shares everything he has, as of right, with his wife
  - Cash transfers are working just fine
- Moral: cash transfers have worked very well around the world, in Africa and Latin America in particular, when they are internally organized, funded, and politically supported
- RCTs that show us that "cash transfers work" are not likely transportable to countries—of which there are many—where governments predate on their citizens

Traveling worms

- Year long war over the internal validity of the Kremer and Miguel paper showing that deworming kids in Kenya increased their test scores
  - See above on internal validity
- Does this study justify "deworm the world"?
  - Seems like a stretch, Nancy and Lucy
  - So when do they work?
- Important "helping factors" that are absent or present
  - Open defection, density of population, toilets or not, soil conditions, worm load, other health conditions
- Again, we have to build models and think about transportation, not jump to global slogans

Graduation RCT, 2015

- Multiple sites: "similar" experiment: Banerjee et al, *Science*, 2015
  - Provision of assets, including cash, advice and counseling
  - Help people escape from a poverty trap
- High quality technique: rerandomization, correction for multiple hypothesis testing, almost perfect acceptance of assignment
  - Similar "effect sizes" across sites: not the same as similar rates of return, for example
- Result is somewhat surprising: wide skepticism about poverty traps
  - So we learn that such things might be possible
  - Proof of concept: efficacy rather than effectiveness in public health
  - As always, without randomization at every stage, cannot tell which component does what (as authors note)

More on graduating from poverty

- To me, it illustrates central difficulty of "what works" agenda
  - All of this is done with NGOs, who have different incentives than the government workers who already don't turn up as teachers, docs, or nurses
  - We already know many things that work: vaccinations, classroom education
  - They DON'T work in practice, because of governance, incentives, principal agent problems: because of *economics*!
  - Does it help to show that something else than won't work works?
- Rajasthan cameras: worked with Seva Mandir, but not in government health service
  - Incentives are a high level force, but the causal mechanism can only operate where social and economic structure permits it
  - The lever on a toaster versus the lever on a toilet

What are RCTs good for?

- We need to think about RCTs as useful tools of discovery
  - Not as magic bullets
  - To use them requires serious economic, social, and political analysis
  - RCTs do not buy us a pass for practicing our professions!
  - It is impossible to do evidence-based policy without knowing stuff
- Nor do they provide a license for encouraging innovations and policies that WE think are desirable
  - Especially without thinking through what is likely to happen when "we" intervene in "their" economic, social, and political system
  - Unintended consequences, many of which develop over time

Ethical questions

How to think about welfare?

Without traditional welfare economics

Beware the anti-politics machine!

- James Ferguson's analysis of an aid program in Lesotho
- Implementers had an analytical framework that was false
  - And was itself politically structured
  - Selected so that it didn't cast a bad light on S Africa during apartheid
- The final beneficiaries of the aid were the thugs who ran the country
- Moral 1: not understanding how the economy and its politics works makes it very hard to help people, and we run the risk of harming them
- Moral 2: what gives US the right to meddle in other societies that we barely understand?
- The RECIPIENTS of altruism are notably absent from much of the EA or RCT literature
  - More than a little paternalism, sometimes explicit, Duflo (2012), Banerjee and Duflo, and a constant risk when behavioral economics tries to think about welfare

Ethical issues in RCTs

- Beware technical solutions to political problems
  - Original sin of modern economic development
- Greenberg and Schroder note that almost all social experiments in the US are done BY rich people ON poor people
  - Before the current wave of development experiments by J-PAL, IPA etc
- NIT experiments designed (by rich people) to improve wellbeing of the poor
  - Used (by other reach people) to minimize the cost of "dealing with" the poor
- Are we using altruism as a cloak to cover using poor people as instruments for our own wellbeing?
  - Without their full, equal and willing partnership, how would we know?
  - Not possible when one side has all the money, Peter Bauer

What's wrong with saving lives?

- Or making people richer? Giving people money?
- It's not so simple

- Surveys in sub-Saharan Africa repeatedly show that improving health is NOT Africans' highest priority (compared with jobs)
  - Yet aid has moved from economic to health aid
- Many anti-poverty schemes are not about saving lives or giving people money
- Even when they are about health and money, there are unintended consequences that can often be exploited by those who have power

Arguments & counterarguments

- Africans may not understand that government CAN improve their health
  - Aid agencies may be better at saving lives than creating jobs
  - And it is OUR money after all, even for effective altruists
- Saving lives: Rwanda, Ethiopia, Goma, Italy
  - Using aid to save lives of children
  - In return for ignoring political abuses? (Rwanda)
  - Many agencies defend such trade-offs: "would you really not vaccinate the kids because you don't like the leader?"
  - How about giving money for child and maternal health when two-thirds is used to train (and arm) their fathers and husbands to commit genocide?
  - Would one third be Ok? How about if genocide is planned for ten years from now?
  - Germans helping Italy drain the Maremma, then bombing it
- Is it possible to improve health from the outside in the long-run?
  - NB none of this argues against basic knowledge

RCT-tested policies

- Policies whose welfare consequences are unclear
  - Incentivizing people to participate in credit schemes
  - Microfinance
- RCTs cannot reveal mechanisms, why people do things
  - Without the why, we can't tell if policy makes them better off
- Used to have revealed preference as our yardstick
  - Basically, that people know what is good for them and act in that light
- Behavioral economics has undermined RP
  - Opening up our ability to tell people what is good for them, with no constraint
  - Back to paternalism

Selective evidence and welfare

- RCTs are not feasible in many circumstances
  - Timing is an issue if policies work differently in the long run than in the short run, as many do
  - So those policies are ruled out if we insist on RCT evidence
- Yet according to traditional welfare economics, it is *exactly* such policies that the state ought to implement, Jeff Hammer
  - Public goods are goods that you cannot provide for yourself
  - And that the market cannot provide
  - Can only be provided collectively
- Benefits of public goods are typically diverse
  - No one can be excluded, nor is it desirable to exclude them
  - That is the point of the public provision
- It is very difficult to evaluate by RCTs such programs
  - This is a serious ethical violation, if we rule out policies that cannot be evaluated by RCTs
  - Throwing away central insights of economics

More selective evidence

- Unintended consequences are typically not included in the outcome measures of RCTs
  - Especially if they happen slowly, or over time
- The provision of a clinic, or cash transfers, might free up governments to buy arms
- Building private schools might lead to hollowing out of government schools
  - Especially if NGOs pay more than the state
- These are difficult or impossible to capture within an RCT
- Yet it is government behavior that is key for poverty reduction
  - We should try to make governments MORE responsive to their citizens, not less

So what SHOULD we do?

- Think seriously about politics, and about the broad consequences of policy change
    - The World Bank and Oxfam, for all of their faults, do this all of the time
    - RCT agenda for all of its virtues, shies away from this
    - Want aid to be freed from the contamination of geopolitics
        - But that is a pure fantasy
- Severe constraint on what can be done from the outside
    - Yet there are policies in donor countries that are hurting people
    - We should stop, and do more of GPGs that help

Conclusions

Measuring and reducing poverty

- We should obsess less over one particular measure of poverty
  - Which plays no role in policy anyway
  - For which no one is politically responsible
  - So even if we fix it, nothing stops the counters straying off the reservation
- We need to think about RCTs as useful tools of discovery
  - Not as magic bullets
  - To use them requires serious economic and political analysis
  - RCTs do not buy as a pass for practicing our professions!