

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

So for the course, we have four days of lectures. Today we'll try to convince you that it was actually a good idea to come here, why randomized evaluation is such a useful tool and why it's superior to many other kinds of impact evaluation.

Once we've convinced you that it's a good idea to come here, then we'll start going through the nuts and bolts of actually how to run randomized evaluations. Tomorrow we'll go over some of the general design possibilities. The following day we'll go into some more of the technical aspects like sample size and measurement. The last day we'll kind of discuss the fact that even in randomized evaluations, things can go wrong, and how you deal with that.

And throughout the entire course as you're learning this, you'll be incorporating what you learn by designing in step with the lectures your own randomized evaluation in the groups that you've been preassigned. So if you check out your name tags, you'll see that you have a group number. Find other people with the same color and number and those will be your group members. We tried to put you together in ways that made sense. So we tried to have people who were interested in agriculture work with other people within agriculture.

And over the course of the four days, you'll be designing your own randomized evaluation. And on the last day, on Saturday, you will be presenting your evaluation to the entire group. So that is pretty much what to expect over the next five days. Have I forgotten anything? Then let me reintroduce Rachel who will start five minutes early.

RACHEL

GLENNERSTER:

The group the Mark was talking about is a really integral part of the course. So unfortunately that is not something that you'll be able to get online. But hopefully it's a chance for you to go through, each time we present an idea in the lecture, you then need to go and apply it in the case that you're developing in your evaluation you're developing in your group. And that's what all the teaching assistants are here, to help you go through the case studies, but also to develop your own evaluations.

So I'm going to start with the general question of what is an impact evaluation and when

should we do one. One of the objectives of this lecture is just to make sure that we're all on the same page when we start using terms like process evaluation and impact evaluation. Because I realize the more I spend time with people who are kind of professional evaluators, I realize that economists and professional evaluators use the same terms but to mean different things, which is incredibly confusing.

So I'm sure a lot of this will be familiar to you, but on the other hand, we need to make sure that every one is at the same level in using the terms in the same way before we kind of head into the nuts and bolts.

But I also have incorporated in this discussion about when you should do an impact evaluation, which is something that comes up an awful lot when I go and talk to organizations who are trying to think through their evaluation strategy. They've heard that they ought to be doing more impact evaluations. There's lots of focus on this. But they're expensive. So how do they decide which of their programs to evaluate and at what stage to evaluate it. They're getting pressure from donors to do this. But they're not quite sure when it's appropriate.

So we'll try and cover those ideas. So we'll start with why is it that we here at J-PAL are focused on impact evaluation. Because there's lots of other things in evaluating our programs that are important. But we just do impact evaluation. We also only do randomized impact evaluation. And that's not to say that's the only thing that's worth doing. We certainly don't think that. That's what we do. There's a reason we do it, because we think it's important. But it's certainly not the only thing that you should be doing in your organizations.

So step back and look at the objectives of evaluation and a model of change, which is very important in terms of how to think about your evaluation, how to design it, different types of evaluation, how evaluation feeds into cost benefit analysis, and then into why to do an impact evaluation, and putting it all together into an evaluation strategy. And then coming back to how do we learn. How do we make an organization that learns from its evaluation strategy rather than just doing this as something a funder wants me to do or I have to do to tick a box. How do I develop an organization that learns from its evaluations and makes it a better organization?

So this is the motivation for what we do. And I think this point is sort of the main point here. If you step back and you think about how much evidence we have in development, to make the decisions that we need to make, it's really quite appalling how little information we have.

If you think about some of the biggest challenges in the world in development about how to

prevent the spread of HIV/AIDS in Sub-Saharan Africa, how to improve the productivity of small farmers across the world, it's really amazing how little really rigorous evidence we have to make those decisions. And we may know that this project may work or that project may work, but we very rarely know what is the most cost-effective place to put a dollar that I have.

If I'm choosing in HIV prevention, if I've got to choose between a lot of different seemingly great projects, what is the project that's going to give me the most bang for my buck? And we really don't have that kind of consistent rigorous impact evaluation data in order to make those decisions. And that was really the reason why J-PAL was started, because of the feeling that we could do so much better if we had that kind of data.

And it's also too often the case that decisions about development are based on emotion rather than data. You can see this in proposals that people write and the discussions that people have, very compelling, personal stories, which are important, but aren't really what we should be making all our decisions on. That may be very motivating to get people involved. But when you're talking about trade-offs, you've got to have a lot more rigorous evidence. If we had that kind of evidence, we could be a lot more effective with the money that we have.

I also think it's true, sometimes people say oh, well you're just talking about passing a dollar and spending it in a slightly more marginally effective way. But what we really need is more money going into poverty really. But arguably, potentially one of the most important ways to get more money to go into poverty relief is to convince people that the money that's going in is actually used effectively. So I don't see these as either/or. Using the money effectively and raising more money, I think, both can come from having more evidence.

So it's also important, I think, in a way, to move from what I think is a very damaging and nonconstructive debate between the aid optimists and the aid pessimists. It's a very kind of polarized debate with Jeff Sachs on one side and Bill Easterly on the other. This is a quote from Jeff Sachs: "I've identified the specific investments that are needed"-- from the previous sentence, you know that this is to end poverty-- "found ways to plan and implement them and show that they can be affordable."

Now if you think we know everything about development already-- we know what's needed, we know how to implement it-- then, kind of, this is the wrong course for you. But I think most of us would agree that that's slightly overstating how much information we have about how to end poverty. There's a lot more questions out there than that suggests. His argument is, but

we have to get people motivated. So we have got to say that we know everything. I don't think we have to be quite that optimistic.

On the other hand, I think this is way too pessimistic. After \$2.3 trillion over five decades, why the desperate needs of the world's poor still so tragically unmet? Isn't it finally time to end the impunity of foreign aid? So Bill Easterly is kind of saying, oh, it has not worked. So let's throw it all away.

We've got to find a middle ground here. It's just about time that we have a development. And they're talking about aid. And I would argue it's much more about development than aid. Aid is only a small fraction of the money that's spent on reducing poverty in developing countries. Development pessimism is just as bad. We've got to think more strategically about not just that all aid is bad or development funding is wasted, but how do we focus the money on the right things.

So that's kind of the motivation for what we're doing. But if you think on a very grand scale, but thinking about the objective of evaluation in general, you can think of it as three things. Accountability, did we do what we say we were going to do? And again, this is true of aid agencies, NGOs, government. And did we have a positive impact on people's lives? So those are two different aspects of accountability that evaluation needs to speak to.

Evaluation isn't only about accountability though. I think it's very importantly about lesson learning so we do better in the future. And that's about does a particular program work or not, and what's the most effective route to achieve a certain outcome? Are there similarities? Are there lessons that you can learn across projects? Are there similarities about what we're finding in the evaluation of this project and that project?

For example, are there are ways that you're learning about how to change people's behavior in health, and agriculture, and education? Are there similarities, sort of underlying principles, that we're learning about that we can use in different contexts? And ultimately, to reduce poverty through more effective programs is the ultimate objective of evaluation.

So using that as a framework, what makes a good evaluation? Well, the key thing is it's got to answer an important question. But it's no good if it answers an important question but it answers it badly. It's got to answer it in an unbiased way. What do I mean by that? I mean that it's got to find the truthful answer to the question.

And really to do that, you need to have a model or a theory of change about how the project is working so that you can test the different steps in the model. And that's the best way to then learn the most. If we simply say, we test whether the project worked or didn't work, we learned something, but we learn an awful lot more if you have a specific model of how the project is going to work and you test the different steps along the way.

Sometimes people say-- this is something that drives me mad in the evaluation literature-- you hear people saying, well, randomized evaluations are a black box. They can tell you whether something works or not, but they can't tell you why. I hope in the next few days, we're going to show to you how you design an evaluation that tells you not just whether it works or not at the end, but why and how, and the steps along the way, and design it cleverly so that you learn as much as you possibly can from the evaluation about the fundamental question. And that's about getting the questions right at the beginning. And it's about doing your model correctly and thinking of indicators along the way that are going to allow you to get to all those steps and really understand the theory of change that's happening.

The model is going to start with what is it we're trying to do, who are the targets, and what are their needs. So in an evaluation in development, this would often be called a needs assessment. And what are their needs? But then what's the program seeking to change? And looking at precise and individual bits of the program, what's the precise program or part of the program that's being evaluated, so asking very specific questions.

So let's look at an example. And all of this we're going to come back to and do in more detail. How do you do a logframe? Again, maybe some of you have done that before. Maybe you haven't. But hopefully you'll learn more about how we think about doing a logframe.

So here's an example of an evaluation that looked, a very simple one, does giving textbooks to children in Kenya improve test scores was the evaluation. But what was the need? What was the problem that the program was trying to test? Well poor children in Busia District in Kenya had low learning levels. They also had low incomes. They had few books. That meant that they couldn't take the books home, and that, the theory was, made it hard to learn. So it was hard to learn because they didn't have a book in front of them in class, but also because there were so few, they couldn't take them home and read up more and do exercises at home.

So what was the input? The input was that a local NGO bought additional textbooks. In order to get to your long-term goal, you not only need the books, you need to make sure that they're

delivered. Because, again, making this chain along the way will help you understand if it doesn't work, where did it go wrong, if the books were bought but they never got there, or they were stuck in the cupboard. How many times have we been to schools and oh, yes, we have lots of books. And we don't want them to get messy when the children are using them. So they're all nicely in their sealed package.

Well you need to be able to distinguish if something doesn't work, is it because it's stuck in the cupboard? Or was it because even when the books are out there they didn't get used or they didn't help? So the books are delivered and used. The children use the books and they're able to study better.

And finally, the impact, which is what we're about here, is yes, you got all of those steps done. But did it actually change their lives? Did it actually achieve the impact you are hoping to get is high test scores? The long-term goal would be not just high test scores but higher income. And that long-term goal may be very difficult to test in the evaluation. And you may use some other work that may have linked these in previous studies in the same country to make the assumption that if we got higher test scores, it will have a positive impact. So again, that's a decision when you make in your evaluation how far along this chain you go, if it's a process evaluation you may stop here. If it's an impact evaluation you have to stop here. But you may not have enough money to take it all the way through to the finest, finest level that you would like to. Oh, I didn't do my little red triangles at right point.

OK. So I've already, in a sense, introduced some of these concepts. But again, let's review them so we know we're talking about the same thing. There are many different kinds of evaluation. And needs assessment is where you go in and look at a population, see what are the issues. How many of them have bed nets? What are test scores at the moment? How many books are there? What's class size? What are the problems in your target population?

In our process evaluation, does someone want to tell me what they would see as a process evaluation? We talked a little bit about it. Someone? Yeah.

AUDIENCE: [UNINTELLIGIBLE PHRASE] the chain that you just presented to see how you get from the input to the output from the output to the outcome of this

RACHEL Right.

GLENNERSTER:

AUDIENCE: Are we successful in doing that in transforming our input into output?

RACHEL Right. So process evaluation looks at did we buy the textbooks? Were they delivered? Were
GLENNERSTER: they used? So moving inputs, outputs, outcomes, but stopping short before we get to the impact. And that's a very useful thing to do, and should be done basically everywhere you do a program, or at least some of those steps need to be measured almost every time you do a program. But it kind of stops short before you get to the impact stage. Have we actually changed people's lives?

We wanted to build a school. Did we build a school? We wanted to build a bridge. Did we build a bridge? We wanted to deliver things? Did we deliver things? But it's stopping before you get the point of knowing whether this has actually changed people's lives.

So an impact evaluation then goes to the next stage and says, given that we have done what we said we're going to do, has that actually changed things? And this is where there was a big gap in terms of what we know. There's a lot of lesson learning you can do from process. But in terms of knowing what kind of project is going to be successful in reducing poverty, you really need to go this next step.

Now we used to just talk about those three. But increasingly, as I say as I have more contact outside economics and the more research side of evaluation to work a lot with DFID and other organizations, other foundations, and other agencies, I realize a lot of what people outside the academic community call evaluation I would call review. It's very weird. Because they would often call what I call something else. But what I mean by review is, it's sort of an assessment. It's sending a knowledgeable person in and reviewing the program and giving their comments on it, which can be extremely helpful if you have a good person going and talking to the people involved, and saying, well, in my experience, it could have been done differently. But it doesn't quite actually do any of these things. It's not just focused in did I build the school. But it's asking questions about was there enough participation. How well organized was the NGO?

And a lot of this is very subjective. So I'm not saying that this is bad. It's just kind of different. And if you have someone very good doing it, it can be very useful. My concern with it, is that it's very subjective to the person who's going. Yeah. Logan?

AUDIENCE: I think that you see so many reviews simply because the way-- you just mentioned DFID. USAID, I think, is the same way. It's all retroactive. The way that contracts are awarded and things like that, usually it's because it's a requirement to evaluate a certain number of

programs. And it's not until after the program is actually done that they decide they're going to evaluate it. And it's obviously cheaper to send one person over and do the simple review.

I think it would be interesting. We'll probably get to this when we talk about how you can apply some of the randomized control test methodology to something that you're doing retroactively.

RACHEL

So just to repeat, the argument is we do a lot of reviews because a lot of evaluation is done

GLENNERSTER:

retroactively. What you can do at that point is very limited. Yes.

So this is a big distinction between the kinds of evaluations, is one that's set up beforehand and one that is after the event. We've got this program. We want to know whether it works. Basically it's really hard to do that. You've already kind of shot yourself in the foot if you haven't set it up beforehand. If we think about what I was saying about it's crucial to have a theory of change, a model about what we're trying to achieve and how we're going to try and achieve it, and measuring each of those steps, if you're coming in afterwards, then you're kind of adhoc-ly making up what your theory of change is. And you haven't set up systems to measure those steps along the way, it's going to be very hard to do. And that's exactly why you end up with a lot of reviews. You're in this mess. And so you just send someone knowledgeable and hope they can figure it out.

To answer your specific question though, you can't do a randomized evaluation after the event. Because the whole point is you're moving people into treatment and control based on a flip of the coin. And then after the event, people have been allocated to the treatment or not the treatment. And it's very difficult to know afterwards were these people similar beforehand. It's impossible to distinguish. They may look different now. But you don't know whether they look different now because they were different in the beginning or because they're different because of the program. Yeah?

AUDIENCE:

I was really interested to read the first case study because it seemed that you were applying randomized control methodology. But it seemed to be actually done retroactively.

RACHEL

No. It wasn't. It might look that way. But it was set up so the first case study uses a lot of

GLENNERSTER:

different methodologies, compares different methodologies. But they couldn't use all those methodologies if they hadn't designed it as a randomized study at the beginning actually. If you've set up a randomized evaluation, you can always do a non-randomized evaluation of it. But if you haven't done it as a randomized to start with, you can't make it randomized.

Prospective evaluation, setting up the evaluation from the beginning, is very important I would say in any methodology you use. But it's impossible to do it. There are a couple of examples where people have done a randomized evaluation afterwards or an evaluation afterwards. And that is because the randomization happened beforehand. But it wasn't done because it was an evaluation.

So if you look at the case on the women's empowerment in India, you will do later in the week, that was not set up as an evaluation. It was set up as a randomized program. And the rationale was they wanted to be fair. So where there's limited resources, sometimes governments are the people who randomized in order to be fair to the different participants.

Some people, in this case, will have to get a women's leader. In Colombia Project, that you'll find on our website, the Colombian government wanted to provide vouchers to go to private school. But they couldn't afford it for every one. So they randomized who would get them.

So that's the one case where you can do a randomized evaluation after the event, when somebody else is randomized beforehand, but they weren't actually thinking of it as an evaluation. But even then, it would've been nice to have data beforehand.

So the last thing on this list is cost-benefit analysis, which is something that you can do with the input from all of these other things. As they say, the piece of information that we have so little of is what's the effect of a dollar here versus a dollar here. And you can only do that if that's one of your ultimate objectives when you're doing these other impact evaluations or these other evaluation methodologies. Because you need to be collecting data about costs. And the benefits will come from your impact evaluation. But you need to get your costs from your process evaluation.

And you can put the two together. And you can do a cost effectiveness. Then if somebody else has done that in their other study, you can do a cost effectiveness comparison across studies. Or even you can evaluate a range of different options on your impact evaluation. And that will give you comparative cost-effectiveness across.

So going into a bit more detail in some of these, needs assessment. We'll look at who's the target population. Is it all children or are we particularly focused on helping the lowest-achieving in the group? What's the nature of the problem being solved? Many of these communities will have lots of problems. So what are we particularly trying to focus on here? Is a test schools? Is it attendance at school? How will textbooks solve the problem? I was talking

about textbooks. Maybe it's because they can take them home. Well, if that's part of your model, your theory of change, you need to be actually measuring that, not just did they arrive.

How does the service fit into the environment? How many times have we sat in an office and designed something that we thought made complete sense, and gone out to the field and thought, what was I thinking? This isn't going to work. How does it feel for the teachers? Do they understand the new books? Do they know how to teach from them? How do the books fit into the curricula?

What are you trying to get out of this? As they say, you want a clear sense of the target population. So then you want to see are the students responding? If you're particularly worried about low-performing kids, are they responding to the textbooks?

Students who are falling behind, a sense of the needs the program will fill, what are the teachers lacking? How are we going to deliver the textbooks? How many textbooks are we going to deliver? And what are the potential barriers for people learning from the textbooks? Then a clear articulation of the program benefits, and a sense of alternatives.

And as I say, if you want to look at cost-effectiveness of alternative approaches, it's very important to think through not just this program in isolation, but what are the alternatives that we could be doing? And how does that fit with them? Is this one is the most expensive things we're going to try, one of the cheapest things we want to try, and everything in between? So you may be thinking in this context, is this a replicable program that I'm going to be able to do elsewhere? Is this the gold-plated version that I'll do if I get lots of funding? Or is this something that I can replicate in lots of other places?

Process evaluation, I've really sort of talked quite a bit about these. So I'm going through them faster. And when you do an impact evaluation, because the impact evaluation is the last thing on that chain, you need to do all the other bits on the chain as well. You can't do an impact evaluation without a process evaluation or you won't understand what the hell your answer is meaning at the end.

So as we say, a process evaluation is asking are the services being delivered? Is the money being spent? Are the textbooks reaching the classroom? Are they being used? And it's also, as I say, important to be asking yourself, what are the alternatives. Could you do this in a better way? Just like a company is always thinking are there ways to reduce costs. You should be thinking are their ways to do this more cheaply.

Are the services reaching the right populations? Which students are taking them home? Is it the ones that I'm targeting or only the most motivated ones? And also, are the clients satisfied? What's their response to the program?

So an impact evaluation, am I missing a top bit here? No. OK. Here we go. We're out of order. So an impact evaluation is, as they say, taking it from there. So assuming once you've got all the processes working, and it's all happening, but if it happens, does it produce an impact?

Take our theory of change seriously. And say what we might expect to change if that theory of change is happening. So we've got this theory of change that says, this is how we expect things to change. These are the processes by which we expect, like the kids taking the books home. So we want to design some intermediate indicators and final outcomes that will trace out that model.

So our primary focus is going to be, did the textbooks cause children to learn more. But we might also be interested in some distributional issues. Not just on average, we might also be interested in was it the high achieving kids that learned more? Was it the low-achieving kids? Because very often in development, we're just as interested in the distributional implications of a project as the average. So who is it who learned?

How does impact differ from process? In the process, we describe what happened. And you can do that from reading documents, interviewing people in administrative records. In an impact question, we need to compare what happened to the people who got the program with what would have happened. This is the fundamental question that Dan is going to hammer on about in his lecture about why do we use randomized evaluations. We talk about this is the counterfactual. What would have happened if the program hadn't happened? That's the fundamental question that we're trying to get at. Obviously it's impossible to know exactly what would have happened if the program hadn't happened. But that's what we're trying to get at. Just one second. Yeah?

AUDIENCE:

So one thing that would seem to fit in somewhere with the impact thing but doesn't quite meet the criteria that you've just described that we've use sometimes is this pre-post test. And that isn't necessarily going to say what would have happened. But it will say, well, what were the conditions when you started? And we extrapolate from that looking at where we are when we ended, what can we say about the impact of the intervention?

RACHEL

Right. So that is one way that people often try and do an impact evaluation and measure are

GLENNERSTER:

they having an impact. And I guess it can give you some sense of whether you're having an impact or flag problems. It's to say, well, what were conditions at the beginning? What are they like now? Then you have this assumption, which is that all the difference between then and now is due to the program. And often that's not a very appropriate assumption. Often things happen.

If we take our example of schools, the kids will know more at the end of the year than they knew at the beginning. Well would they have known more even if we hadn't given them more textbooks. Probably. So that's kind of the fundamental assumption you're making. And it's a difficult one to make.

It's also the case that we talked to people who were doing a project in Gujarat. And they were tearing their hair out and saying, well, we seem to be doing terribly. Our program is doing terribly. People now are worse off than when we started. This was, well Mark will know the years of the riots and earthquake in Gujarat.

They'd basically taken data when they started. In the meantime, there had been a massive earthquake and massive ethnic riots against Muslims in Gujarat. Of course people were worse off. And that's not because of you. So it can go either way actually. You can assume that your program is doing much better because other things are coming along and helping people. And you're attributing all the change to your program.

Or it could be the case in this extreme example. There's a massive earthquake and massive religious and ethnic riots and you attribute all the negative to your program. So it's a way that sometimes people use of trying to get an impact. It's not a very accurate way of getting your impact, which is why a randomized evaluation would help.

So, as you say, it doesn't quite fit this criteria. Because it doesn't quite answer. It says what happened over the period. It doesn't say what would have happened. It is not a comparison of what would have happened with what actually had. And that's how you want to get at your impact. So there are various ways to get at it. But some of them are more effective than others.

So let's go back to our objectives and see if we get a match these different kinds of evaluations to our different objectives for evaluation and find out which evaluation will answer which question. So accountability: the first question for accountability is just did we do what we

said we were going to do. Now that you can use a process evaluation to do that. Because did I do what I said I was going to do. I promised to deliver books. Did I actually deliver books? Process evaluation is fine for that kind of level of accountability.

If my accountability is not just did I do what I said, but did what I do help? Ultimately I'm there to help people. Am I actually helping people? That's a deeper level of accountability. And that, you can only answer with an impact evaluation. Did I actually make the change that I wanted to happen?

If we look at lesson learning, the first kind of lesson learning is, does a particular program work or not work. So an impact evaluation can tell you whether a particular program worked. If you look at different impact evaluations of different programs, you can start saying which ones worked, whether they work in different situations, or whether a particular kind of program works in different situations or not.

Now what is the most effective route for achieving a certain outcome is kind of an even deeper level of learning. What kind of thing is the best thing to do in this situation? And there you want to have a cost-benefit analysis comparing several programs based on a number of different impact evaluations.

And then we said an even deeper level is, can I understand how we change behavior? Understand deep parameters of what makes a successful program, of how do we change behavior from health to agriculture? What are some similarities and understanding of how people tick, and how we can use that to design better programs? And again, that's linking our results back to theories. You have got to have a deeper theory understanding it, and then test that with different impact evaluations. And you can get some kind of general lessons from looking across impact evaluations.

And then if we want to have our reduced poverty through more effective programs, which is our ultimate objective of doing evaluations, we've got to say, did we learn from our impact evaluations? Because if we don't learn from them and change our programs as a result, then we're not going to achieve that.

And I guess to say that solid, reliable impact evaluations are a building block. You're not going to get everything out of one impact evaluation. But if you build up enough, you can generate the general lessons that you need to do that.

I've said quite a lot of this already. But needs assessments give you the metric for defining the cost-benefit ratio. So when we're looking at cost-benefit analysis, we're looking at what's the most cost-effective way of achieving x? Well, you need a needs assessment to say what's the x? What's the thing that I should be really trying to solve? Process evaluation gives you the costs for your inputs to do a cost-benefit analysis. And an impact evaluation tells you the benefit. So all of these different inputs and needed to be able to do an effective cost-benefit analysis.

AUDIENCE: Rachel?

RACHEL Yeah?

GLENNERSTER:

AUDIENCE: The needs assessment seems to be more of a program design sort of a [UNINTELLIGIBLE], whereas the remaining three are more like the program has already been designed and we are being cautious, we have thought that this is the right program to go with. Please design a process evaluation for this or a program evaluation for this. How has that needs assessment different from the one that feeds into program design?

RACHEL Well, in a sense, there's two different concepts here. You're right. There's a needs

GLENNERSTER: assessment for a particular project. We're working with an NGO in India called [UNINTELLIGIBLE] working in rural Rajasthan. And they said, we want to do more on health in our communities. We've done a lot of education and community building. But we want to do a lot more in health.

But before we start, we want to know what are the health problems in this community? It doesn't make sense to design the project until you know. So we went in and did., what are the health problems? What's the level of services? Who are they getting their health from? We did a very comprehensive analysis of the issues.

And that was a needs assessment for that particular NGO in that particular area. But you can kind of think of that in a wider context of saying, what are the key problems in health in India or in developing countries? What are the top priority things that we should be focusing on? Because again-- and I'm going to get on to strategy in a minute-- if you're thinking as an organization, you can't do an impact evaluation for everything. You can't look at comparative cost-effectiveness for outcomes in the world. Or at least you've got to start somewhere. You've got to start on, what do I most want to know? What's the main thing I want to change to see

what's the cost of changing that thing?

So is it test scores in schools? Or is it attendance? Am I most concerned about improving attendance? If you look at the Millennium Development Goals, in the sense, that's the world's prioritizing. They're saying, these are the things that I most want to change in the world. And there they made the decision, rightly or wrongly, on education, that they wanted to get kids in school. And there isn't anything about actually learning.

And whether your needs is getting kids in school or learning, you would design very different projects. But you would also design different impact evaluations, because those are two very different questions. So the needs assessment is telling you what are the problems? What am I prioritizing for my programs, but also for my impact evaluations? Yeah?

AUDIENCE: Do you need to make a decision early on whether you're interested in actually doing a cost-effectiveness analysis as opposed to a cost-benefit.

RACHEL [INTERPOSING VOICE]

GLENNERSTER:

AUDIENCE: [UNINTELLIGIBLE PHRASE]. efficiency measure, where as cost [UNINTELLIGIBLE]--

RACHEL So I'm kind of using those too easily interchangeably. I don't think it's so important here. How

GLENNERSTER: would you define the difference between them?

AUDIENCE: As I understand it, but [UNINTELLIGIBLE PHRASE] they getting a better answer. But cost-effectiveness is a productivity measure. And it would mean that you would have to, in an evaluation say, OK, I'm going to look at I put one buck into this program and I get how many more days of schooling out of it. Right?

RACHEL Right.

GLENNERSTER:

AUDIENCE: Whereas cost-benefit requires that it all be in dollars or some other [UNINTELLIGIBLE].

RACHEL So you've got to change your benefit into dollars. So I'll give you an example of the difference.

GLENNERSTER:

AUDIENCE: Like [INAUDIBLE PHRASE].

RACHEL

Let's make sure everybody's following this discussion. A cost-effectiveness question would be to say, I want to increase the number of kids in school. How much would it cost to get an additional year of schooling from all of these different programs? And I'm just assuming that getting kids in school is a good thing to do. Right? I want to do it. So I'm asking what's the cost per additional year of schooling from conditional cash transfer, from making it cheaper to go to school by giving free school uniforms, or providing school meals. There are many different things I could do that will encourage children to come to school. But I know I want children to come to school. I'm not questioning that goal. So I just want to know the cost of getting a child in school.

Cost-benefit kind of squishes is it all. And it really asks the question, is it worth getting kids in school? Because then you can say, if I get kids in school, they will earn more and that will generate income. So if I put a dollar in, am I going to get more than a dollar out at the end? I'm not going to flick all the way back to it. But if you remember that chart that went through the process and impact, and then the final thing of high test scores was higher income, ultimately am I getting more money out of it than I'm putting in?

I think that's sort of a philosophical decision for the organization to make. It's very convincing to be able to say, for every dollar we put in, I think for the deworming case that you've got and you do later in the week, they do both, cost-effectiveness and cost-benefit. And the cost-effectiveness says, this is the most cost-effective way to get children in school.

But they also then go further and say, assuming that these studies that look at children in school in Kenya earn higher incomes are correct, then given how much it costs to get an additional year of schooling, and given an assumption about how much extra kids will in the future because they went to school, then for every dollar we put in, I think it's you get \$30 back.

So you kind of have to make an awful lot more assumptions. You have to go to that final thing and put everything on income. Now if I was doing the women's empowerment study, then I'm not sure that I would want to reduce women's empowerment to dollars. I might just care about it. I might care that women are more empowered whether or not it actually leads to higher incomes.

So it kind of depends on the argument that you're making. If you want to try and make a this is really worth it, this is a great program, not just because it's more effective than another

program, but that it generates more income than I'm putting in. That's a great motivation. But I wouldn't say you always have to reduce it to dollars. Because you have to make an awful lot of assumptions. And we don't necessarily always want to reduce everything to dollars.

So here it is. We've just been talking about it. So this is a cost-effectiveness. So this is the cost per additional year of schooling induced. We're not linking it back to dollars we're measuring. We're just assuming that we want kids in school. Millennium Development Goals have it as a goal. We just think it's a good thing, whether or not it generates income.

What we did is take all the randomized impact evaluations that had as an outcome getting more children in school and calculated the cost per additional year of schooling that resulted. So you see a very wide range of different things.

Now conditional cash transfers turn out to be by far the most expensive way of getting an additional year of schooling. Now that's partly because mainly they're done in Latin America where enrollment rates are already very high. So it's often more expensive to get the last kid in school than the 50th percentile kid in school. And then the other thing, of course, in general, things cost more in Mexico than in Kenya, especially when you're talking about people. Teacher's wages or wages outside of school are more expensive.

But the thing that was amazing was that providing children with deworming tablets was just unbelievably cost-effective. So \$3.50 for an additional year of schooling induced. And putting it this way I think really brought out that difference.

The other thing I should say in comparing this is, there were other benefits to these programs. So Progreso actually gave people cash as well. So it wasn't just about getting kids in school. So of course it was expensive, right? And we haven't calculated in those costs. In cost-benefit, if we reduced everything to dollars, it would look very different because you've got a value of all these other benefits.

But again, deworming had other benefits. It had health benefits as well as education benefits. So we're just looking at one measure of outcomes here.

AUDIENCE: Excuse me.

RACHEL Yeah?

GLENNERSTER:

AUDIENCE: Are these being adjusted for Purchasing Power Parity, PPP?

RACHEL So this is not PPP. This is absolute. So again, we've sort of debated it backwards and

GLENNERSTER: forwards. So if you're a country, you care more about PPP. But if you're a donor and you're wondering whether to send a dollar or a pound to Mexico or Kenya, you don't care about PPP. You care about where your dollar is going to get most kids in school. So there's different ways of thinking about it.

It sort of depends on the question you're asking and who's the person. For a donor, I think this is the relevant way. If you're a donor who only cares about getting kids in school, this is what you care about. We also can redo this taking out the transfers. There's this other benefit, the families, of getting the money. So this is the cost to a donor. So that's one way of presenting it. But you can present it in other ways too.

AUDIENCE: Can you also sometimes do a cost-benefit of the evaluation itself?

RACHEL That's kind of hard to do because the benefits may come ten years later. The way to think

GLENNERSTER: about that is to think about who's going to use it, and only do it if you think it's going to actually have some benefits in terms of being used and not just maybe within the organization. But if it's expensive, is it going to be useful for other people? Is it answering a general question that lots of people will find useful?

So often evaluations are expensive in the context of a particular program. But they're answering a question the lots of other people will benefit from. So the Progresa evaluation has spurred not just the expansion of Progresa in Mexico, but it has spurred it in many other countries as well because it did prove very effective. Although it's slightly less cost-effective in these terms. But it led to an awful lot of learning in many, many other countries. So, I think, in that sense, it was an extremely effective program evaluation.

AUDIENCE: Excuse me, I just have a question on that very last item there.

RACHEL OK, so this one is even cheaper and it's a relatively new result. But it only works in certain

GLENNERSTER: circumstances. When people don't know the benefits of staying on in school, ie., how much higher wages they're going to get if they have a primary education, then telling them that information is very cheap. And both in the Dominican Republic and Madagascar-- so two completely different contexts, different rates of staying on in school, different continents, very different schooling systems-- in both cases it was extremely effective at increasing the number

of kids staying on in school. But that only works if people are underestimating the returns of staying on in school.

If they're overestimating them, then it would reduce staying on in school or if they know already, then it's not going to be effective. So this is something that I think is a very interesting thing to do, and again, is worth doing. But you need to first go in and test whether people know what the benefits of staying on in school are.

Basically they just told them what's the wage if you complete primary education versus what's the wage if you don't complete primary education. It's very cheap. So if it changes anything, it's incredibly effective.

AUDIENCE: Is the issue of marginal returns a problem? Do you have to say that every program is only relevant to places where it's at same level of enrollment or admission?

RACHEL
GLENNERSTER: Well this is a sort of wider question of external validity. When we do a randomized evaluation, we look at what's the impact of a project in that situation. Now at least you know whether it worked in that situation, which is better than not really knowing whether it worked in that situation. Then you've got to make a decision about whether you think that is useful to another situation. A great way of doing that is to test it in a couple of different places.

So again, this was tested in two very different situations. The deworming had very similar effects. In rural primary schools in Kenya, it works through reducing anemia. Reducing anemia in preschool urban India had almost identical effects. Getting rid of worms in a non-randomized evaluation to be true, but kind of a really nicely designed one in the south of the United States had almost exactly the same effect.

So they got rid of hookworm in the 1900s. And again, it would increase school attendance, increase test scores, and actually increase wages just from getting rid of hookworm. And they reckoned a quite substantial percentage. This paper by Hoyt Bleakley at Chicago found that quite a substantial difference in the income of the North and the South of United States in 1900 was simply due to hookworm. So this is being tested.

So ideally you test something in very different environments. But you also think about whether it makes sense that it replicates. So if I take the findings of the women's empowerment study in India where it works through local governance bodies that are quite active in India and have quite a lot of power, and tried to replicate in Bangladesh where there is no equivalent system, I

would worry about it. Whereas worms cause anemia around the world. And anemia causes you to be tired. And being tired is likely to affect you going to school. That's something that seems like it would replicate. So you have to think through these things and ideally test them.

If I'm doing microfinance, would I assume it has identical effects in Africa, or Asia, and Latin American? No. Because it's very dependent on what are the learning opportunities in those environments. And they're likely to be very different. So I'd want to test it in those different environments.

We're falling a bit behind. So I promised to do when to do an impact evaluation. So there are important questions you need to know the answer to. So that might be because there's a program that you do in lots of places, and you have no idea whether it works. That would be a reason to do one.

You're very uncertain about which strategy to use to solve a problem. Or there are key questions that underline a lot of your programs, for example, adding beneficiary control, having some participatory element to your program. It might be something that you do in lots of different programs when you don't know what's the best way to do it or whether it's being effective. An opportunity to do it is when you're rolling out a big new program. And you're going to invest an awful lot of money in this program, you want to know whether it works.

This is a tricky one. You're developing a new program and you want to scale it up. At what point in that process should you do the impact evaluation? Well you don't want to do it once you've scaled it up for everywhere. Because then you find out it doesn't work, and you've just spend millions of dollars scaling it up. Well that's not a good idea.

On the other hand, you don't want to do it when it's your very first designs. Because often it changes an awful lot in the first couple of years as your tweaking it, and developing it, and understanding how to make it work on the ground. So you want to wait until you've got the basic kinks ironed out. But you want to do it before you scale it up too far.

We've done a lot of work with this NGO in India called Pratham. And we started doing some work for them. And by the time we finished doing an evaluation, their program had completely changed. So we kind of did another one. So we probably did that one a little bit too early. But on the other hand, now they're scaling up massively. And it would be silly to wait until they'd done the whole of India before we evaluated it.

AUDIENCE: You said it may be more appropriate to do a process evaluation initially to get a program to the point where it can be fully implemented and all the kinks are worked out.

RACHEL Yeah, exactly. If we're going back to our textbook example again, you don't want to be doing it

GLENNERSTER: until you've got your delivery system for the textbooks worked out, and you've made sure you've got the right textbook. It's a bit of a waste of money until you've got those things. And exactly, a process evaluation can tell you whether you've got those things working.

The other thing that makes it a good time or a good program to do an impact evaluation of is one that's representative and not gold-plated. Because if Millennium Development Villages, \$1 million per village. If we find that that has an impact on people's lives, that's great. But what do we do with that? We can't give \$1 million to every village in Africa.

So it's not quite, what's the point? But it's less useful than testing something that you could replicate across the whole of Africa, that you have enough money to replicate in a big scale. So that's interesting because you can use it more. Because if you throw everything at a community, yes, you can probably change things. But what are you learning from it?

So it takes time, and expertise, and money to do it right. So it's very important to think about when you're going to do it and designing the right evaluation to answer the right question that you're going to learn from.

AUDIENCE: If a program hasn't been successful, have you found that the NGO's have abandoned that program?

RACHEL Yes, mainly. We worked with an NGO in Kenya that didn't work. They just moved on to

GLENNERSTER: something else. Pratham, we actually did two things, both of which worked, but one which was more cost-effective than the other. And they dumped the computer assisted learning even though it was like phenomenally successful. But the other one was even cheaper.

So they really scaled that up. And they haven't really done computer assisted learning even though it had a very big effect on math test scores. And compared to anybody else doing education, it was very cost-effective. But compared to their other approach, which was even more cost-effective, they were like, OK. We'll do the one that's most cost-effective.

Now there are some organizations that kind of do one thing. And it's much harder for them to stop doing that one thing if you find it doesn't work. They tend to think, well, how can I adapt it? But these organizations that do many things are often very happy to, OK, that didn't work.

We'll go this direction.

So we want to develop an evaluation strategy to help us prioritize what evaluations to do when. So the first thing to do is step back and ask, what are the key questions for your organization? What are the things that I really, really need to know? What are the things that would make me be more successful, that I'm spending lots of money on but I don't know the answer, or some of these more fundamental questions, as they say, about how do I get beneficiary control across my different programs.

The other key thing is you're not going to be able to answer all of them by your own impact evaluations. And as they say, it's expensive to do them. So the first thing to do is to go out and see if somebody else has done a really good impact evaluation that's relevant to you to answer your questions already. Or half answer or more gives you the hypotheses to look at.

How many can I answer just from improved process? Because if my problems are about logistics, and getting things to people, and getting cooperation from people, then I can get that from process evaluation. So from that you can select your top priority questions for an impact evaluation and establish a plan for answering them. So then you've got to look for opportunities where you can develop an impact evaluation that will enable you to answer those questions. So am I rolling out a new program in a new area? And I can do an impact evaluation there.

Or you might even want to say, I want to set up an experimental site. I don't really know whether to go this way or that way. So I'm just going to take a place and try different things. And it's not going to be really part of my general rollout But I'm going to focus in on the questions. Should I be charging for this or not? How much should I charge? Or how should I present this to people? And you can take a site and kind of try a bunch of different things against each other, figure out your design, really hone it down, and then roll that out. So those are two kinds of different options of thinking about how to do it.

And then, when you've got those key questions of your impact, you can combine that with process evaluations to get your global impact. What do I mean by that? Let's go back to our textbook example. If you're giving out textbooks across many states or throughout the country, you've evaluated it carefully in one region. And you find that the impact on test scores is whatever it is. And then you know very carefully, and maybe you've tested it in two different locations in the country and you've got very similar results.

So then you can say, well I know that every time I give a textbook, I get this impact on test schools. Then from the process evaluation, you know how many textbooks are getting in the hands of kids. Then you can combine the two, multiply up your impact numbers by the number of textbooks you give out.

Malaria control with bed nets, if I hand out this many bed nets, then I'm saving this many lives. I've done that through a careful impact evaluation. And then all I need to do is just count the number of bed nets that are getting to people and I know my overall impact. So that's a way that you can combine the two. You don't have to do an impact evaluation for every single bed net you hand out. Because you've really got the underlying evaluation impact model, and you can extrapolate.

AUDIENCE: Rachel?

RACHEL Yeah?

GLENNERSTER:

AUDIENCE: Do you think in the beginning when you got a program that you're interested in, do you think that's the moment to think about the size of the impact that you're looking at that people expect? And also, as part of that, what's going to be the audience, the ultimate audience that you're trying to get to if you're successful with a scale-up. And those two things, I think, frequently come together. Because it's the scaling up process where people are going to start to look at those cost-effectiveness measures and cost-benefit.

RACHEL

GLENNERSTER: I mean, I would argue that you've always got to be thinking about your ultimate plans for scaling it up when you're designing the project. Because you design a project very differently if you're just trying to treat a small area than if you're thinking about, if I get this right, I want to do it on a much wider area. If you've always got that in mind, you're thinking a lot about is this scalable? Am I using a resource that is either money or expertise that is in very short supply, in which case there's no point in designing it this way because I won't be able to scale it beyond this small study area.

So if that's your ultimate objective, you need to be putting that into the impact evaluation from the moment. Because there's no point in doing the impact evaluation, very resource-intensive project, and say, well, that works. But I can't do that everywhere. Well then what have you learned? You want to be testing the thing that ultimately you're going to be able to bring everywhere.

So in a lot of our cases, we're encouraging our partners to scale it back. Because you won't be able to do this on a big scale. So scale it back to what you would actually be doing if you're trying to do the whole of India or the whole of this state. Because that's what's useful to learn. And you want to be able to sell to someone to finance the scale-up. So I think having those ideas in your mind at the beginning is very important, and as they say, making it into a strategy, not a project by project evaluation, but thinking about where do I want to go as an organization. What's the evidence I need to get there, and then designing the impact evaluations to get you that evidence.

And people often ask me about how do you make sure that people use the evidence from impact evaluations. And I think the main answer to that is ask the right question. Because it's not about browbeating people to make them read studies afterwards. If you find the answer to an interesting question, it'll take off like wildfire. It will be used. But if you answer a stupid question, then nobody is going to want to read your results.

So we're learning from an impact evaluation, so learning from in a single study did the program work in this context? Should we expand it to a similar population? Learning from an accumulation of studies, which is what we want to get to eventually, is did the same program work in a range of different contexts, India, Kenya, south of the United States? And that's incredibly valuable because then your learning is much wider and you can take it to many more places.

Did some variation in the same program work differently, ie., take one program and try different variants of it and test it out so that we know how to design it.

Did this same mechanism seem to be present in different areas? So there's a lot of studies looking at the impact of user fees in education and health. You seem to get some very similar results. And again, that's even more useful. Because then you're not just talking about moving deworming to another country. You're talking about user fees. What have we learned about user fees across a lot of different sectors? There's some common understandings and learnings to take to even a sector that we may not have studied before.

And then, as they say, putting these learnings in the place, in filling in an overall strategy of what were my gaps in knowledge? And am I slowly filling them in? So, I think that's it.

So I'm sorry the last bit was a little bit rushed. The idea was to kind of motivate why we're

doing all of this. Today you're going to be in your groups. The task for your groups today, as well as doing the case, is to decide on a question for an evaluation that you're going to design over the next five days.

So hopefully that's made you think about what's an interesting question. What should we be testing? Because I think often an underlooked element of designing an evaluation is what's the question that we want to be answering with this evaluation? Is it a useful question? How am I going to use it? What's it going to tell me for making my program, my whole organization more effective in the future? So any questions?

AUDIENCE: What would you say are some of the main limitations of randomization? So I assume one of them is extrapolate the populations that are different? Are there other main ones that you can think of?

RACHEL
GLENNERSTER: So it's important to distinguish when we talk about limitations. One is just general, what's the limitation to say, extrapolating beyond? But the other thing is to think of it in the context of what's the limitation versus other mechanisms?

Because, for example, the extrapolating to other populations is not really a limitation of randomized evaluations compared to any other impact. Any impact evaluation is done on a particular population. And so there's always a question as to whether it generalizes to another population. And the way to deal with that is to design it in a way that you learn as much as you possibly can about the mechanisms, about the routes through which it worked.

And then you can ask yourself when you bring it to another population, do those routes seem like they might be applicable, or is there an obvious gap? This worked through the local organization. But if that organization isn't there, is there another organization that it could work through there or not?

If you think that deworming works through the mechanism of anemia, well it works between the mechanism of there being worms and of anemia, you can go out and test. Are there worms in the area? Is the population anemic because there may be worms and they're not anemic.

So that's a way to design the evaluation to limit that limitation or reduce the problem of that limitation. But it's not like the very active flipping a coin and randomizing causes the problem of not being able to extend it. It's true of any impact evaluation.

I think one limitation which you will find in your frustration as you want to try and answer every single question that you have, and you get into the mechanics of sample size and how much sample size do I need-- and again, that's not necessarily just of a randomized evaluation, but any quantitative evaluation-- you can test a limited number of hypotheses. And every hypothesis you want to test needs more sample. And so the number of questions you can answer very rigorously is limited. And I think that's the limitation that we often find very binding. Again, any rigorous quantitative evaluation will have that limitation.

We'll talk a lot tomorrow about sometimes you just can't randomize. Freedom of the press is not something that you can randomize except by country. And then we'd need every country in the world. It's just not going to happen. So we'll look at a lot of new techniques or different techniques that you can use to bring randomization to areas where you think it would be impossible to bring it to.

Compared to other quantitative evaluations, you sometimes have political constraints about where you can randomize. But as I say, quantitative versus qualitative, the qualitative isn't so limited by sample size constraints. And you're not so limited to answer very specific hypotheses.

The flip side is you don't answer any specific hypotheses. And it's the same rigorous way. But it's much more open. So very often what we do is we combine a qualitative and quantitative, and spend a lot of time doing qualitative before to hone our hypotheses, and then use a randomized impact evaluation to test those specific.

But if you sit in your office and design your hypotheses without any going out into the field, you will almost certainly waste your money because you won't have asked the right question. You won't have designed it. So you need some element of qualitative to make sure some needs assessment, some work on the ground to make sure that you are asking, you're designing your hypotheses correctly because you've only got a few shots. Yeah?

AUDIENCE: I was wondering, do you know of some good evaluation or randomized impact evaluation on conservation programs?

RACHEL On conservation programs? I can't think of any, I'm afraid, but eminently doable. But we can

GLENNERSTER: talk about that if you can persuade your group to think about designing one. Anyone else think of a conservation program? Yes?

AUDIENCE: I don't have an example. I wish I did. And you've mentioned this. I just need to really underline it for myself. A lot of the programs that my organization does are comprehensive in nature. So they have lots of different elements meant to in the end, collectively, [UNINTELLIGIBLE PHRASE]. What I'm understanding here is that you could do an impact evaluation of all of those collectively. But really it would be more useful to pull them out and look at the different interventions side by side or something. Because that way you'll get a more targeted--

RACHEL
GLENNERSTER: It's true. The question was, if you have a big package of programs that does lots of things, you can do an evaluation of the whole package and see whether it works as a package. But in terms of learning about how you should design future programs, you would probably learn more by trying to tease out, take one away, or try them separately. Because there might be elements of the package that are very expensive but are not generating as much benefit as they are cost. And you would get more effect by doing a smaller package in more places. You don't know unless you take the package apart.

Now then if you test each one individually, that's a very expensive process. Because it needs a lot of sample size to test each individually. There's also a very interesting hypothesis that's true in lots of different areas. People often feel, where there are lots of barriers, so we have to attack all of them. It only makes sense. You won't get any movement unless you do. There are lots of things stopping kids going to school. There's stopping, say, girls going to school. They're needed a home. There are attitudes. There is their own health. There's maybe they are sick a lot. So we have to address all of those if we're going to have an impact.

We don't know, the answer is. And indeed, in that example where we're working with Save the Children in Bangladesh, they had this comprehensive approach. Where there are all these problems. So let's tackle them all. We convinced them to divide it up a bit, and test different things, and see some of their own worked, or whether you needed to do all of them together before you changed anything, which is a perfectly possible hypotheses and one that a lot of people have, but hasn't really been tested. The idea that you've got to get over a critical threshold. And you've got to build up to it. And only once you're over there do you see any movement.

Well actually on girls going to school, it's quite interesting. Most of the evaluations that have looked at, just generally, improving attendance at school, have had their biggest impact on girls. I should say most of those were not done in the toughest environments for girls going to school. They're not in Afghanistan or somewhere where it's particularly difficult. But it is

interesting. Just general things and approaches in Africa and India have had their biggest impacts on girls, which suggests that you've got a hit every possible thing is maybe not right. Yeah?

AUDIENCE: [INAUDIBLE PHRASE] and the political constraints [INAUDIBLE PHRASE].

RACHEL Right. So we'll talk actually tomorrow quite a lot about the politics of introducing an evaluation
GLENNERSTER: or at least the different ways that you can introduce randomization to make it more politically acceptable. That's slightly different from whether the senior political figures want to know whether the program works or are willing to fund an evaluation.

I've actually been amazingly surprised. Obviously we find that some places. There are certain partners or people we've started talking with. And you can see the moment the penny drops that they're not going to have any control. Because you're going to do a treatment comparison. You're going to stand back. At the end of the day the results going to be that. There's no fiddling with it, which is one of the beauties of the design. But it will be what it will be, which is kind of why it's convincing. But there are certain groups who kind of figure that out. And they run for the exit because there's going to be an MIT stamp of approval evaluation potentially saying their program doesn't work. That's life. Some people don't want to know.

The best thing I can say in that situation is test alternatives. It's much less threatening to test alternatives. Because there's always some alternative of this versus that, that people don't know. And then you're not raising the does it work. You're saying well, does this work better than that? And that is much less threatening. It doesn't tell you quite as much. But it's much less threatening.

There's a report called *When Will We Ever Learn*, looking at the politics of why don't we have more impact evaluations, which was very pessimistic. But if you look at somewhere like the World Bank that just put a purse of money for doing randomized impact evaluations out there. And anybody in the bank could apply. And people were like why? There's no incentives for them to do it. Program office's, they've already got a lot on their plate. Why would they add doing this? It's going to find out that they're opening themselves to all these risks because maybe their program doesn't work.

Massively oversubscribed, first year, six times more applicants than there was money. It just came out of the woodwork as soon as there was some money to do it. So I'm not saying every organization is like that. Obviously not everybody in their bank did that. But it was, to me,

actually quite surprising how many people were willing to come forward.

Now we have the luxury of working with the willing, which if you're working within an organization, you don't necessarily have that luxury. You will see as you get into the details of these things, that you need absolutely full cooperation and complete dedication on the part of the practitioners who were doing these evaluations alongside the evaluators.

You can't do this with a partner who doesn't want to be evaluated. It just doesn't work. They are so able to throw monkey wrenches in there if they don't want to find out the answer. Then it's just not worth doing it because it's a partnership like that. It's not someone coming along afterwards and interviewing. It is the practitioners and the evaluators working hand in hand throughout the whole process. And therefore if the practitioners don't want to be evaluated, there's not a hope in hell of getting a result.

We should wrap up. A lot of these things we're going to talk about. But I'll take one more.
Yeah?

AUDIENCE: How important, or how relevant is it, or how much skepticism can there be about a case where the evaluators and the practitioners work for the same people or are funded by the same people?

RACHEL
GLENNERSTER: Yeah, we've even got practitioners as co-authors on our studies. This is another place where I kind of part company from the classic evaluation guidelines, which say that it's very important to be independent. I'd argue if your methodology is independent, what you want is not independence. You want objectivity. And the methodology of a randomized evaluation can provide you the objectivity. And therefore you don't have to worry about independence.

Now there's one caveat to that. The beauty of the design is you set it up, as I say. Well you don't stand back in the sense. You've got to manage all your threats and things. But you can't fiddle very much with it at the end. The one exception to that is that you can look at subgroups.

So there was an evaluation in UK of a welfare program. And it was a randomized evaluation. And there was some complaining. Because at the end, they went through and looked at every ethnic minority. and then you know I can't remember whether it did work in general. But it didn't work for one minority, or it didn't work. But anyway, you can find one subgroup for whom the result was flipped. And that was the thing on the front page of the newspapers, rather than

the overall effect.

So there's a way to deal with that, which is increasingly being stressed by people who are kind of looking over the shoulder and making sure that what is done in randomized evaluations is done properly, which is to say that you need to set out in advance-- we'll talk about this a bit later on-- but you need to set out in advance what you're going to do.

So if you want to look at a subgroup like does it affect the lowest performing kids in the school differently from the highest performing kids-- do I care most about the lowest performing kid-- if you want to do that, you need to say you're going to do that before you actually look at the numbers. Because even with a randomized evaluation, you can data mine to some extent. Well if I look at least ten kids, does it work for them? If I look at least ten kids, does it work for them?

Statistically you will be able to find some subset of your sample for whom it does work. So you can't just keep trying 100 different subgroups. Because eventually it will work for one of them. So on the whole, you need to look at the main average effect. What's the average effect for the whole sample? If you are particularly interested in a special group within the whole sample, you need to say that before you start looking at the data.

So that's the only way in which you get to fiddle with the results. And otherwise it provides an enormous amount of objectivity in the methodology. And therefore, you don't have to worry so much about a Chinese wall between the evaluators and the practitioners, which, I think, is incredibly important. Because we couldn't do the work that we do if we had that Chinese wall. It just wouldn't make sense, doing your theory of change, finding out how it's working, designing it so it asks the right questions. None of that would be possible if you had wall between you. So it just wouldn't be anything like as useful. So getting your objectivity from the methodology allows you to be very integrated with the evaluation, and practitioners to be very integrated.