

ANNOUNCER: 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.

PROFESSOR: Great, hi everyone. It's great to be here, and I hope you guys are having fun so far. It's great to meet all of you too. So this is the day where we talk about how to randomize. A few other topics will come up here and there, and hopefully you will stop me whenever you have questions, and let me know. So don't hesitate to stop even if it's just a clarifying point or a deeper, more substantive question about what I'm saying.

The three basic components of this morning's lecture will be first about methods of randomization, and that'll be-- the majority of what we'll talk about is going through a few different ways that we talk about randomization when we talk randomization. One of the common misperceptions that I've seen in the world when I am meeting with organizations or governments for the first time, and they've heard of a randomized trial as a concept, they often have a very well defined and narrowly defined concept of what that means.

And the fact is, there's a lot of creative ways that we go about doing randomized trials that adapt to different settings, because there's a lot of situations in which you can't do what might be considered the most standard prescription drug type randomized trial. And so we have to be a little bit creative in settings and understanding, what are the constraints we're facing, and how can we adapt the methodology to fit in this particular setting? Or maybe not, right? So that's going to be methods of randomization, we'll go through a few of the key approaches that we use.

The second, and this is a topic which really-- to say this is topic one, topic two, topic three isn't quite exactly right. And by the time we finish one, we're going to have talked a lot about number two and number three, a lot of the number two being gathering support for evaluation. And the point here is that one of the reasons why we choose one method over another when we're thinking about how to go about setting up the design is because some methods are going to be easier for gathering support for evaluation, and so that's part of a back and forth process with organizations and situations. And then we're going to try to walk through a typical plan, so to speak.

Perhaps in my mind, one of the single most important things to remember about doing an evaluation is to remember that we're not trying to just ask, how did we do? There's nothing wrong with asking that, but it's very short-sighted. What we should be asking is, what should we do? And that's the point of a good evaluation, is to guide us in future decisions.

If you're a donor, and you're running a huge initiative, and you're spending \$20 million or something-- I'm just picking a number-- and you want to do an evaluation of this. But because of whatever the nature of your program is, this is it, this is the only time you're ever going to do it, and it's a weird program that you believe in, but it's weird, and no one else is ever going to do it. I realize it's kind of a weird example, just go with me.

And so that's a situation in which, I think, most reasonable people would say, why are you doing an evaluation? What's the point? Is it just to pat yourself on the back? Is that really the goal? Because if there's not future money that's at stake, future money that we have to decide how are we going to spend, what's the point of doing all this, other than just to see whether I made a good decision in the past or not? But that's not really useful, that's not why we're here. We're here because we realize that there are tons of future decisions being made, and we need better information in order to make those decisions. And we need those as donors, but we also need those as organizations.

And that's one of the key things going back to point number two in the outline. How do you get organizations on board and excited and involved in evaluation? It's when the evaluation is actually able to speak to questions that they have. And so good evaluations often help to identify the key implementer's questions and answer them. And I know I'm making that sound really simple, like oh, that's all we have to do. But the key really is to taking that type of approach when working with organizations. How do you turn these things into win win for operations?

If you're a leader in an organization, and you're not a researcher, so you kind of understand the value of research, and it sounds like a good thing, but you're hired because you need to go and deliver these services, and you need to be efficient in delivering your services. You want to know that the research is nice and good and needs to be done, but isn't going to get in your way. Or if it is going to get in your way, you're going to get something for it.

And that's a very common attitude, and I can respect that attitude, if someone is just really focused on operations. And so the question is, how can we design research, how can we listen

to what the operations people are saying about what their challenges are, what their struggles are, the choices that they're making that are tough? And actually build into the research ways of helping them answer those questions. So that's something that we often aim to do.

Methods of randomization. I'm going to walk through four different methods that we will often use: basic lottery, a phase in, a rotation, and encouragement. And we'll talk about each of these. These are not all mutually exclusive methods, to be clear. So I'm just going to point out there's different ways and levers for doing things. Is that readable? OK.

Sorry, let me skip that slide, and we'll come back to it at the end to recap. So let's start with the simplest, which is a lottery. A lottery is like a clinical trial, where if we were running a test for prescription drugs, it would be a very standard, regimented process. We'd be perhaps in some hospital. We'd have some sort of intake process with people who are in the hospital and have certain criteria, and then we'd approach them and say, there's a new drug. It's experimental. There's risks, there's potential rewards, it's up to you. You have the following disease, these are the issues you're facing. What do you want to do?

But that's a situation where we would take 1,000 of the people. They would all get informed consent being told about what the risks were. There would be parameters set up so that if the outcomes were proving very decisive one way or another, the study would end. But it's fairly straightforward from a statistical standpoint and from a research design standpoint. You bring them in, 1,000 are entered into the study, 500 of them are randomly chosen to get a pill, 500 are randomly chosen to get a placebo, and you measure the outcomes, whatever they may be, from getting the pill versus the placebo.

And so the question is, can we apply this in social science, outside the laboratory type setting? So some of the constraints that we face when we try to take that really simplistic way of doing things. So the first is that we often can't-- when you're doing a randomized trial on a prescription drug, the whole point is the research study. That is what it is. There is no program around this. It's a study to see what the effectiveness of a particular pill is.

If we're working with an organization that is trying to do teacher training in schools, trying to issue loans, trying to promote savings, trying to teach agricultural practices, there's a program there that's involved. And you can't just go in and change things around in the program without paying attention to what this actually means for the goals of the program has.

So the second really important element is that it must be perceived as fair. Now the thing that's

the single most common situation that I find myself in is when I'm dealing with organizations that are capacity constrained. They only have enough money to go to 200 schools. They only have enough money to make 1,000 loans. They only have enough marketing people to visit 5,000 households and promote savings, or whatever the activity is they're doing, there's some capacity constraint.

And so one principle that I've actually taught my children, very young, they all know-- so like last week, we had to fly, and I hadn't seen them for the weekend. So the kids all wanted to sit next to me. I have three kids. And it was resolved very simply, no whining, no nothing. We just randomly chose one. And they all knew immediately. Now in fairness, I've done this before. But they knew instinctfully just that this is-- there's no complaining here. There's no favoritism. We took each of their boarding passes, we flipped them upside down. We chose one.

Actually I said it backwards. It was one that couldn't sit with me. It was two with me and one without, so we chose the one who didn't. Now ironically, the two that did win ended up both falling asleep immediately, and so I switched seats in the end. But that's a different issue.

But the point hopefully is clear, that there's really, in some respect, nothing more fair than a random process. It is giving everybody who is eligible, who's aware or not aware, depending on how the set up is, but everybody who is within a setting of-- who has access to a program, and giving them all the same chance of participating.

Now in a lot of settings, we would actually suggest that this is actually more fair than, for instance, letting politics and letting nepotism and letting any sort of other favoritism play in to deciding who gets things and who doesn't. Which we all know, in many places are serious issues for the allocation of any sort of resource, that depending on how it happens, then that could actually be the worst possible outcome that we would want to see as philanthropists, as utilitarian individuals interested first and foremost in alleviating poverty.

But you could tell other stories in which there is processes where you really want to reach the very poorest of the poor. And so you have to-- let's come back to that issue towards the end. I'll use that as an example later in the lecture.

So it must be politically feasible. Now, it might not be politically feasible if people who have the power in the local settings are not willing to do it, because they want to be able to choose the people in their networks to provide the services to them. So the politically feasible could be for the exact reason that we want to do it randomly. It also can just be politically infeasible for

other interpersonal reasons. The right people who need to be part of the decision making process just haven't really brought into the value of an evaluation.

And there are situations we face all the time like that, where there's nothing we can do. There's just someone who is the decision maker who needs to be on board with something, and is simply not on board with doing a rigorous evaluation, and perhaps concerned about the results that are going to come out of the project, in terms of is suspicious of external evaluators. They feel like we know what works and what doesn't, and they don't want to relax control to an outsider.

There's also obviously situations where people just feel like what do I have to lose-- I'm sorry, what do I have to gain? Where organizations will be of the ilk that, look, we have lots of media, lots of attention for what we're doing. Everybody tells us it's a great idea. We're executing, we're implementing. What do I have to gain from doing an evaluation. If we absorbed more money right now, I wouldn't know how to spend it. And there are organizations that I've interacted with that are more or less of that ilk. And so there's clearly situations like that.

Must be ethical. So this is something we're always very concerned about and cautious about. And there are many situations we find ourselves for the reasons I mentioned above, in terms of fairness, that the random process is arguably the more ethical process. But they're clearly situations that someone could put forward where you could say, well, wait a second, that's not good. That's not right. So these are issues that one has to be sensitive to.

I think one of the things that's most important here is not necessarily whether something's-- we might all be able to analytically agree on the ethics, but that doesn't mean that everybody else will agree and perceive things the way we perceive them. And so even if we can analytically understand that a random process is fair, if someone is just bringing a certain bias to the table in terms of the way they think about a random processes, then this can be a problem. And again, this is one of persuasion and personality more so than logic and philosophy.

So, why resource constraints are an evaluators best friend. I think I've actually now said most of this, but basically, most programs have limited resources. And so examples of where this has been done is training programs for entrepreneurs or farmers. School vouchers is perhaps one of the single most common examples we've seen this done, where there's a government program to provide school vouchers for private school or secondary school, or even college, and they just can't educate the entire population of people who want to go to private school.

And so there's a voucher program, there's an enrollment process, you apply. There's a public lottery literally done openly through the newspapers or on TV.

And that type of transparency is done in order to make it politically feasible, so that everybody can see. If it's not done that way, then the concerns become about whether is really truly random, it was done behind closed doors. Was the politician really just kind of pulling out his favorite people? So lotteries become, in the simplest of case, it's often the starting point. If we can pull off a public lottery in this way, or private, either way, then it really does become the simplest approach. So when it's possible, it's nice. But there are clearly situations that come up where this is not going to be possible.

So there's also flexible ways of doing the lottery. So let's go through a few different scenarios here. So first of all, there's often a question about the unit over which you randomize. What we mean by this is, if you're going to do a lottery, do you do a lottery at the individual level? Or do you do a lottery, for instance, at the village, by offering an entire village a package of services? Or if you're building wells, for instance, this is actually a treatment that's really being done at the village, not the individual. There might be some individuals who live closer to the well than others further, but you're really choosing a village and building wells.

And so the lottery system is flexible in this way. The general concept is the same. We've done interventions in Ghana, for instance, where the randomization was done at the village level for receiving a community driven development type program. And so different geographic clusters were all put into a public lottery, and half of them randomly chosen to receive financial resources in order to help build an epicenter, and lots of labor in terms of helping and guiding and facilitating the process for building these quote, epicenters.

So the second point is that sometimes when we do lotteries, sometimes it's kind of a full versus partial. What we mean by this is whether it's done all at once in one kind of public way, or whether it's done through an ongoing process. And depending on how things are set up, sometimes it's done one way or the other. So the school vouchers is an example where we would do it all at the very same time. There's a whole bunch of people who apply for a school voucher program. There's 5,000 applicants, there's 2,000 school vouchers, we randomly choose 2,000, they get the school vouchers.

In a situation that we've done in South Africa and also the Philippines, when we build it into a credit scoring process. And so every day, there's 10 applications for a loan. The bank is not

going to issue all the loans. They take those 10, they put them in three piles, yes, no and maybe. And then what happens is they take the maybes and they say, we only want to make half of these. And then we randomize which half of those get credit and which do not. And this is something we've done now a few times in order to measure the impact of credit.

From the bank's perspective, this is exactly one of the cases I was talking about before where there's a win win from an operations standpoint. This is a bank who says, these are genuine maybes. We don't know if we should be lending to them or not. We're not sure if it's profitable for us, we're a bank. So this is a method for them that helps mitigate their risks in deciding what their portfolio should look like as a whole. And from our perspective, provides this nice lottery system where some people are randomly assigned to get credit and others not.

So let's go back, and now let's think about some of the things that will often happen when you do a lottery design. Suppose you have 500 applicants and you have 500 slots. At first glance, you might think you're kind of screwed. You set up this nice, big process, and lo and behold, it turns out you didn't have over-subscription. You thought you did. You thought you were going to use this over-subscription to randomize who gets in and who does not. So what can you do in this type of situation? So there's some low hanging fruit type answers, and there's also a possibility that this wouldn't work.

But the first is, could you increase the outreach activities? That a lot of times in this situation, what this really means is that whatever was done to market this program was not effective in getting in the right number of people. Because the intent was to get 1,000 or 2,000 people applying, and only 500 applied. So now think about, from an operations standpoint, what does that tell you about applying?

And that might be one of those cases where there's actually useful, interesting research for the organization. If the organization was saying, we're going to get 2,000 people applying for this, and then only get 500, well, it tells you that maybe you could do something to help them learn how they can get their marketing up. And so then you can test out different approaches for marketing that helps the operations learn more about what is it that brings in people to apply for these scholarships, or whatever it is that's being done, and at the same time, benefits the program from increasing-- I'm sorry, benefits the evaluation from getting a larger intake.

The risk with this is that you end up bringing people into the program who weren't really part of the target program. So if that's the answer, then this is actually a really bad idea to go out and

do more extensive marketing. If you have to go through extra leaps and bounds in order to bring people in, so much so that it changes the nature of what the program is, that's a situation where you might want to go back to the drawing board and think again about what the right answer is.

Suppose there's 2,000 applicants, and suppose in the process of doing this that the organization that's doing things says, you know what? There's 500 worthy candidates and there's 500 slots. So a simple lottery would not work. So we have some sort of screening process, and we ranked them. And we can rank them, and we put numbers by everyone. So why should we do anything other than just taking the top 500?

So when that type of thing happens, a lot of times the questions that we want to ask then are about what was the screening process that was being used here? So let's go back to the credit scoring study that I was referring to. The thing that we're struck by when we've done this credit scoring is really how wide that maybe category is, when you really get into the nuts and bolts in talking with the lenders about where their data are coming from, and the quality of the data, and how they're individual judgment weighs in to influence where people fall within those three buckets, the yes, no, maybe. And the fact is that the profitability of the person at the high end of the maybe is really not much different, if at all, from the low end of the maybe.

And so the point is that when you get inside and figure out, what is it that's really going on that made them say, well, we had 2,000 applicants and we had 500 eligible, so we're done. If you got inside the box a little bit more and talked with them about how they got to those 500, things would start coming out of that process that might be the exact areas where you can say, why is that necessary as a criteria? Is that really something you want to filter on to bring people in or out? Yeah?

AUDIENCE:

Let's say your criteria are, one is gender and one is educational level for the purpose here. And let's say that, for whatever reason, you can't be sure that the educational level is really what they say it is. So you could go ahead and select on the basis of educational level, knowing that maybe people are not telling the truth and that would be OK. You could remove that as a criteria entirely, or you could find some other kind of proxy. That might work, right?

PROFESSOR:

Right.

AUDIENCE:

OK. Kind of random.

PROFESSOR: But that's exactly kind of one of the key things that would come out. So the point is, let's just go with the middle example. If you really think that the education was what you wanted to screen on, but you don't have confidence in what you're looking at as actually being a reliable measure of education. But yet that's causing a filter to draw you down to the 500. But then when you get inside and you realize this is a really bad measure of education, why are we using this? And then all of the sudden you relax that one rule and you're up to 900 people, it tells you that maybe this isn't such a good way to be filtering, and I should be doing the 900. And while I'm at it, maybe I should try to think about how to measure education better or something.

So let's take another kind of example, which is-- and this actually mimics the story I have up here about training, but the credit scoring example is another example exactly of this. Where you end up-- and I said that right here, sorry-- you can think about having two different piles of people that, when you say there's 500 that are eligible, really maybe what you have is not 500 that are eligible, but you had 200 that are really eligible. You really want them in the program, they must be there. And then the next 300, you're a little more questionable on. And the next 300 really weren't so different from the 300 after those.

And so what you can do in that type of situation is set up a process where you say look, if you have certain eligibility requirements, you're in. And then you're also not part of the evaluation. And it's the next 300 that we're going to combine with the following 300 after that, we're going to put together a pool of 600, and we're going to randomize those.

Now this has a clear benefit and a clear cost. The benefit is that you can now get a very nice, clean estimate on the impact of those 600. The cost is we've changed the research question here. We've changed the evaluation question. We're no longer answering the question, what is the impact on everyone who receives this service? And that's not a good thing. We don't want to lead with methodology and then force fit questions. We want to set the research questions. We have to ask ourselves, how much are we losing by only studying those individuals? And in some settings, those are the exact individuals you want to study. But in some, maybe that's not so the case.

So in the credit scoring, I think of those as the exact people we want to study. Because when we think about programs that expand access to credit, what we're doing is we're talking about those people on the bubble, and we're talking about ways of getting them access that they didn't have otherwise. And the people who have really, really good credit scores and are very

credit worthy, they're not the ones we're thinking about when we think about expanding access to credit.

So let me give you another example of one where this would be bad. So we are doing these programs about targeting the ultra-poor, where we go into countries-- I'm sorry, we do go into countries, but we go into villages-- and we first identify the 20 poorest people in the village. And then of those 20, we hold a lottery, and 10 receive services and 10 do not. And then we do this in about 30 villages. And the organizations we're working with, it's exactly the setting I'm describing here. The organizations we're working with have a fixed budget. They can provide services in each situation to 1,200 people, and that's it.

And if we went into 120 villages, they can do 10 per village. If we went into 60 villages, they could do 20. But either way they look at it, they got 1,200 people that they're going to be able to provide these services to. It's going to be an asset transfer, they're going to be on goats and training and consumption bundles.

Now, if in some these villages people said, well, wait a second. We have the 20 poorest people in the village, yes. But four of these people are really just standout poor. They're much, much poorer than everybody else. So we want to exclude the four very, very, very poorest and make sure that they get it with certainty, and then only evaluate the other 16. So this would actually be a bad thing from the evaluation perspective. It might be the right thing to do, and we can talk about what the trade offs are. We're not actually doing that, and I'll explain why. But from an evaluation perspective, that would be now changing the research question in a bad way, as compared to the credit scoring, where I was arguing that it's a good change. So why? Well, because this is a problem where we really do actually want to know the impact on those bottom four as well as the next 16.

And so if we did something where we just allowed the very bottom four to get the program with certainty, and then only evaluated the next 16, we're missing an important part of the sample frame. Now, the reason why, in these settings, we've done it as the full 20 is because realistically, when we've actually gone into these villages and done this type of exercise, it's difficult, if not impossible, to get consensus that there's really four people that stand out.

And the fact is, even when you're measuring poverty-- and we do have some objective things, we all know that there's lots of components to poverty. There's lots of ways of measuring it. It's not even just consumption this month, its vulnerability in general as a concept, which could be

about the variation over time, and how vulnerable you are, and how your social networks are for helping you absorb shocks. There's just many different ways of measuring it, and it's unrealistic to think then we can go into a village and really draw these amazingly fine lines to say these people are stand out different than the rest. When you're going into a village of this size and you're finding the 20 poorest, those 20 are statistically indistinguishable from each other, more or less. So that's the philosophy of the organizations. And that's the reason for doing it the way we're doing it.

So sometimes, exclusion is not desirable. Sometimes there's no way, whatever the context is, of excluding individuals. So other points that we will often use is the expansion of a program, or any sort of program where there's some sort of initial stage, and then what you're doing is you're simply randomizing where that initial stage is, or where the program expands into. A program is only going to go to so many villages, and they can't exclude, but they can control where their individuals, where their credit officers, where their program officers, education trainers, whatever it is that they're doing, they obviously can control what villages they go to and what they do not.

So when there is a process like this, where there's gradual expansion, we often will think about doing a program evaluation using a phase in approach. A phase in approach basically says, look, we're going to go to all 1,200 of these villages over the next three years, but we can randomize which ones we go to in each year. So everybody's going to get services in the long run.

Now one thing that's nice about that is, particularly when there's community led interventions, there's oftentimes a desire to have some involvement from everybody in the program at some point in time in the program. And this is often something that organizations do ask for. So phase in approaches allow for that more naturally, because everybody's going to be receiving a service at some point.

Similarly, the rotation-- which I'll mention in a second-- is a very similar twist on the phase in. Where the rotation, instead of having it be where you slowly phase in a process, it's a process where everybody's actually receiving a service at all points in time, and we're just randomizing who gets what. So we'll talk about some examples in a moment about when that can work.

So some key advantages is that everyone gets something eventually, and this provides incentives to maintain contact as well with control villages. They're not just participating in

surveys, although we do do a lot of surveys sometimes where there is no intervention related, and people are often more than-- at least in our experience-- more than happy to participate in this interesting, weird thing, with these people coming and asking us all these questions. But having said that, there are situations where you want that continuous support and continuous buy in to what we're doing, and so the advantage is that it provides them some incentive to maintain contact.

Some of the concerns is that it does complicate estimating long line effects. If the goal was to study something over 10 years, but everybody got phased in at the end of three, well, you can't study the 10 year effects. You can study the effect of getting something for two more years 10 years later, but that's not nearly as interesting as asking what the 10 year effect is from getting a particular service. So let me give you an example of a situation where the first is actually perfectly interesting.

Take neonatal care. If you want to study the effect of neonatal care, doing a phase in across villages is perfectly fine, because this is something that is going to be affecting infants. And so when you provide that neonatal care, and you do this as a phase in, you're now studying those children, and this is perfectly fine. You can study now the effect of the neonatal care over a 10, 15 year horizon, even though it was phased in to everybody. Because once it's phased in, the control area of those kids are three, four, five years old, and so it doesn't apply.

So I think the main thing to talk about in terms of phase ins that does become an issue about expectations. So let me give you the simplest example. In the world of credit, this is usually something that concerns me a lot when we're talking about doing studies. And in fact, I've personally been involved in studies where there was a proposal to do a phased in credit program and we said no. We didn't do it, because we were very concerned about what happens to a control group individual who was told, you're going to get a loan, but just please wait six months, or year, or year and a half, or two years, whatever the amount is.

Now the longer you have to wait, the less it's an issue. But if it's a relatively short period of time, like a year, then the question is, well, what were they going to do with that money? And is it something they're willing to wait a year for? And if the answer is yes, well then this is a real problem for thinking that this is a valid control group, because what it says is, we're going to see a delay in an activity specifically because of getting access to this loan in one year. So what it says is they did have other options. They were a little bit more expensive perhaps, a little more costly either in time or interest. And they said, you know what? I don't need to build

that new roof now. I don't need to buy that new sewing machine for my enterprise. I'll go ahead and just put it off, and I'll do it in a year.

And what it would do is it would lead us to overestimate the impact of getting access to credit, whereas if they weren't promised this good loan in the year, they would've just borrowed at a little bit of a higher interest rate right now. And they still would have made the investment in their business, they just would have had a higher interest cost that would cause us to overestimate the impact of our program, because we would only-- the true impact is just really a savings in interest, not about access to credit in a binary sense, but just about the price.

But yet what we would then find ourselves measuring is seeing a treatment group of people who fix their homes and bought sewing machines for their businesses, and a control that didn't, and we would think that aha, there was a real binary constraint here where people were actually held back from getting access to credit. And this program didn't have that positive effect.

So a rotation design. So a rotation design is basically groups getting treatment in turns. Group A gets treatment in the first period, group B gets treatment in the second. The main advantage is it's perceived as fair and easier to get accepted. Everybody's getting something, we're just randomizing what you get in a given round. So we have the same anticipation issue that we just mentioned with phase in as one concern here. It depends on what the two treatments are, but if you really want that other treatment that other people are getting, and you're told you're going to get it in a year, this could affect your behavior now for the same reason.

Also it does have the same long term problem of the phase in, in that everybody-- if we're just rotating-- everybody is getting the treatment. Now, another twist on rotation is not rotating just-- call it like a placebo design, so to speak. Suppose that you just have two different treatments doing two totally different things, everybody gets something. And you use one to measure the impact of the other and vice versa. So it's similar to rotation, except not going full circle. It just means everybody gets something.

So in something like that, you don't have this long term impact issue, or this issue. But you do have a problem where it's not so clear what you're comparing anymore. So you'd have to be really careful and think about what it is you're trying to do. We tend to think that a lot of interventions have indirect effects in many facets of our life.

So if you're providing training about entrepreneurship to one group, and another group you're

providing some health service, and you think, well, this is great, because what does entrepreneurship training have to do with health? And what does health have to do with entrepreneurship training? So I can provide my health services over here and measure the health outcomes for my other group and compare them. And the same thing with business activity.

But it's not too hard to tell stories where these are going to interact with each other. So you're healthier, and this makes you more able to work, and your business does better. Your business does better, this makes you richer, and you spend more money on health. It's not hard to tell stories across two seemingly unrelated sectors where you will have that type of effect on each. You have to think about those types of issues. Yeah?

AUDIENCE:

Can a rotation work well in agriculture? A given country would have-- so you'd like to give the farmers-- everybody gets something. Some people get credit, some people get seeds, some people get a variety of things. But obviously some places have good soil, some places the farmers are near a road, and the data that you need to look at your sampling design is kind of tricky, because you may not know that much about all the variables that would impact a farmer's success. Rotation seems a nice way to-- because there are so many different treatments that people can get, that it would seem pretty tricky to implement if there were--?

PROFESSOR:

So I think, in that type of setting what you're describing, we'll come to one in the end. But let me just say an example here, which is the question you're proposing about agriculture is about how different treatments will interact with other treatments, and with underlying context. So there's two things going on in your question. One is how does soil quality affect whether a certain treatment is effective or not? And the second is maybe credit alone is bad, and maybe seeds alone is bad, but the two together is good, and things of this nature.

So that's not a setting where we would think instinctively about a rotation design. That's a study where we would think about two things. One is making sure that our study is being done in a wide enough variety of soil, to use that example, so that we can actually study the effect on one soil quality and another. And then the second thing is we were thinking about having multiple treatments, but not in a rotation style, but in a way that you have some people, they get seed. Some people, they get training. Some people, they get seed plus training. Some people, they get nothing. So we'll get to an example like that hopefully towards the end, but that's that design.

A rotation design is really more about when you're doing something that realistically will not have an indirect effect on the other group. So the example I'm going to give you is a rotation study that was done, the Balsakhi Case that's done, I think it's one of the cases in your reading. It's what what you do this morning, right? So that's a classic rotation example, because what we're doing is, some schools got third grade, and some schools got fourth grade, and then they rotate. And the idea is as long as the third graders don't affect the fourth, and the fourth graders don't affect the third, then this is good, and every school got something. And then we just rotate around what they're getting. And that's what we mean more by rotation design.

The key here is that this is a great example of where the rotation design was a useful way of getting the support of the schools. It's going to the schools and get them to agree to do all these tests with the children. And it would have been hard to just go in and get them to do tests without being offered some service along with that. Now they're willing to accept that the service only went to one grade, not both. They understood it was a phase in within their school, one gets it one year, that other, the next. But that was the way of getting the schools to cooperate was by offering it through this type of rotation design.

So next is encouragement design. Now encouragement designs are-- first of all, this is orthogonal to everything else I've been saying. Encouragement design can be done on top of a phase in, on top of a rotation, on top of a lottery. This is not mutually exclusive with the others that I've discussed. Yeah?

AUDIENCE: I just had a question on the phase in approach.

PROFESSOR: Yeah?

AUDIENCE: So suppose you wanted to roll out packaging group one, and then you'll roll out packaging group two. When you roll out packaging, you notice that something is not working that well, and then you wanted to tweak it a little bit, you wanted to change the spec. And for the sake of the experiment, are you not supposed to tweak it when you roll it out to the second group, you just keep it the same? Because [UNINTELLIGIBLE PHRASE]?

PROFESSOR: Right. So great question. I think the key here is to think about the timeline. I don't really have a chalkboard. The key is to remember that with the phase in-- so let's go with a really simple phase in, two waves. So in that setting, the second group, when you do the treatment with them, that's actually after the study is over. So the idea is that they're really your control group,

but they're participating with you because they know they're going to get it in the future, or whatever the circumstance is.

So that's a situation in which the answer is yeah, you can do whatever you want with them. But if you did know from operational observations that the treatment itself wasn't working so well, then just remember that when you're evaluating something, what you're evaluating was a program which you already think from operational reasons was less than effective. And so that should perhaps inform you a little bit about things like what to measure in terms of what you want to put in the followup surveys.

I suppose we could complicate your question a little bit and add a third wave. So you have three waves, one for each year. And after the first year you learn, oh, turns out we shouldn't have done it like this. We should have done it differently, and so you want to change things for the second wave. And that's perfectly fine. It does mean now when you're doing your analysis, you should think about this as two studies. You have your first wave, and you can compare that to your wave three, that is control for the entire study. And then you have your second wave, and you can look at them for one year and compare them to wave three, and you really have two different studies in that setting.

So encouragement designs, like I said, this is not mutually exclusive to the others. And encouragement design, just think about what the word means. It means we're encouraging people to do something. We're not forcing, we're not mandating. That means the control group does not necessarily have nobody getting services, and a treatment group does not necessarily have everybody getting the service. There's simply something done to encourage people to do something, to participate.

Now the key here is to think about what we're really saying, is the control in the phrase randomized control trial, the reason for the word control is this idea that the researcher has some control over the process in deciding who gets a service and who doesn't. So now we're just moving the control, and it's no longer over who gets the service and who doesn't. It's over who's offered the service and who's not, or who has some encouragement to get the service or not. And we still have perfect control if it's executed properly over that offer, over that suggestion, that encouragement. But we don't have perfect control over who actually gets the service.

So a very simple example of this is suppose that I gave each of you a marketing brochure to

go to Au Bon Pain during lunch and go because of their delicious scones, and I only gave it to half of you. I am now encouraging half of you to go, the other half not. Anybody can go to Au Bon Pain, I'm not controlling that. And if I wanted to then, for some reason, study the effect of going to Au Bon Pain, I could do that. I'm not sure what the point of that would be. But the point is, I'm only controlling who receives this offer and who doesn't. I'm not controlling who actually goes to Au Bon Pain and who does not.

And so this is often the easiest thing to control in the process. And the entire key here from a statistical perspective-- not the entire, we'll go into some other issues-- is about what that differential usage rate will be among those who were encouraged and those who were not. And when you get into power calculations later in this week, that's going to be a very important element to think about. Because if that encouragement is really, really weak and just barely changes people's behavior, it means you need a huge sample.

In an extreme, an encouragement design is exactly a perfectly controlled randomized controlled trial. An encouragement that gets people who get the marketing, every single one of you goes to Au Bon Pain, and if you didn't receive the marketing, nobody goes. Statistically it's the same now as a perfect lottery system. But usually when we're doing encouragement design is when we have some expectation for it not to be perfect, and so we're using that.

So what makes something a good encouragement? So I think there's two things to think about that are important. One is that it's not itself a treatment. The minute the encouragement design itself becomes a treatment, then we have to think about what is it that you're actually evaluating here. Your goal is for your encouragement to be totally innocuous, to just by chance, by randomness, some people will be more likely to use a service than others. So you want it to be as innocuous as possible.

So a good idea is typically marketing. We typically think of marketing as a good approach, just making people aware of a service makes them more likely to use it. So we've done marketing experiments, for instance, in the Philippines a lot where we're doing some sort of door to door marketing of a savings product offering people savings. Anybody in the village could walk into the bank and open a bank account. But realistically, only those who get a knock on their door become aware enough of it to actually go and open up a bank account.

Here's a bad idea. Let's provide training to people that encourages them to use credit. So let's bring them in, let's give them a big course about business management and how to use credit

in order to take out a loan. And let's use that as an encouragement tool for measuring the effect of credit, because after doing this course, they'll be more likely to borrow. So the problem with this, if we want to look at business outcomes, is we just gave them a month long course in management of an enterprise. And that alone is going to have an impact on their enterprise, we think, we hope. And so if it does, well then, what are you measuring the impact on? Was it an impact of the training program? Or was it an impact of getting access to credit? And you can't separate these out at all.

So this first thing to think about is just making sure that treatment is really innocuous. In econometrics, we refer to this as the exclusion restriction, in that what it's saying is that we want to make sure that the only-- if we're going to draw a link from the encouragement to the take up decision to the outcome measure we care about, that the only path through which the encouragement affects the outcome is as it generates higher take up. If it has its own effect outside of the decision to take up, now it's a problem econometrically, and we can't really claim that we're measuring the impact of using the service. We could only measure the net effect of the two together.

So the second issue is for whom are we estimating the treatment. So here's my favorite tongue in cheek example for this. Suppose we went into a village and we offered free alcohol to anybody who takes out a loan. Might be great in the first stage in the sense that it generates lots of higher borrowing. But what are we measuring here in terms of who we're studying? We're studying people who respond to this particular incentive of free alcohol. That's certainly not the program that we're typically trying to evaluate when we're trying to do an evaluation of microcredit. And we want to make sure that we're getting the people in the study who are the right people, who are the types of people that are thought about as the target audience for a microcredit program. That means not drunkards.

And so you want to make sure, you do want whatever that approach is to be something that is sensible, that seems somewhat in the scope of normal. Or at least doesn't create a sample selection bias in the sense that it doesn't make the people who take up the program be particularly different in a way that is not useful. Yeah?

AUDIENCE:

So an example that I'm thinking about is access or information about a microcredit program to participants who are typically very uninformed about these things. So the information is out there. Theoretically, it's really accessible. But we know that unless we tell them that this program is out there for them, chances are very good that they would never think of it on their

own. So that would not then be a good situation for this kind of a thing, because we know that we are in effect offering them a sort of special in by the very effect of offering it, even though theoretically it's accessible.

PROFESSOR: I would actually says that's actually a perfect setting. To do this, let me rephrase the question, which is suppose you have a program, and only the highly informed normally are going to come in. And so in order to do an encouragement design, what you're doing is you're going out and you're only going to move the people who are not highly informed. The highly informed already know about you. They're either coming in or they're not. You give them information, it doesn't matter, I already knew about this. So what you're doing is you're moving the less informed people, you're informing them about the service you're offering, and now they're coming in or not as they wish. But they're more likely to come in now than the people who are not informed.

So this is a perfectly relevant approach if it's the case that this is an organization that does aspire to grow, and they're going to grow through informing people. In most of the settings I've been involved in, at least the type of information we're dealing with is usually not much different than what they do normally. It's just marketing. It's just targeted and controlled marketing, where we control what villages they go to do the marketing, or what household's doors they knock on.

But in a lot of situations, the encouragement design literally has them doing exactly the same operation's that they normally would do. But it's just recognizing that it's still a voluntary decision. They can't make someone borrow. They're going to a village, they're holding meetings, they're presenting what they do, and some borrow and some don't.

AUDIENCE: I think I'm saying something slightly different, but that might [UNINTELLIGIBLE PHRASE]. So among the group who would not normally know about this, it's not that I'm going to-- I'm not saying the group who knows, forget about them. I don't know, I'm confusing myself. We're assuming that the group of people that we work with to provide our program, we would provide a precursor program, and we would say among people that we work with, half of them we would tell, and half of them we won't tell. Is that what you're saying too?

PROFESSOR: You would go out of your way to give them information about the program. Everyone can get in, but you go out of your way to approach half and tell them about the services.

AUDIENCE: Understanding that chances are that if we don't tell them, they won't go, because they're just

uninformed? OK.

PROFESSOR: Right. Yeah?

AUDIENCE: I don't want to interrupt if there was more to this exchange. My question is about distinguishing thing between marketing and training. If the treatment is something like a financial product or service that's poorly understood, and you don't want to-- do you think that maybe financial literacy is an important determinant, but you want to isolate just access to the product or service and keep financial liberty separate? How do you draw the distinction between marketing and training?

PROFESSOR: I can tell you in one setting, here's what we did to try to understand this better. Let me restate the question, which is how do you distinguish between marketing and training? This is really a spectrum. So an example I gave that was bad was a month long training program, and I said it's fine to just knock on a door. Why am I drawing the line there? And it's a perfect question, and I can tell you that the first time I actually ever did this type of design, we actually had an entire treatment group that was just knocking on doors, but with no product. It was just to test out whether the knocking on the door had an effect on savings.

So we had two treatment groups. We had a treatment group that got a commitment savings account. A bank officer went to the door, knocked on it, gave them a pitch about why they need to save, and savings is good, and here's a goal. You should have a goal for savings. And here's an account that we'll offer you to help you reach your goal. It wasn't a very long pitch, but it was a marketing visit. And we had a pure control that got no contact from the bank.

And then we had a second treatment group that we called the marketing treatment group. And this group got the knock on the door, got the same pitch for about 5, 10 minutes about why it's important to save, and how the bank is there to help them save, but didn't get offered that special savings account that had special rules to it. And that's done exactly to try to understand where to draw that line. So if it's a situation where you're particularly concerned, then you could actually think about having treatments designed specifically to test whether there is a direct effect without the treatment you really care about. Yeah?

AUDIENCE: I'm kind of jumping onto the last point as well. But the one encouragement design that I'm familiar with is one where a subsidy was actually utilized, but then it was a random distribution of who was offered the subsidy. And, for that matter, because they're trying to determine a

demand curve, the subsidy varied. So maybe you'd be offered 35% off, maybe you'd be offered 75% off. Still random in who was given the offer, but then they had the encouragement to take off based on how big the subsidy was.

But it seem my initial-- not knowing enough about the details of the program-- is that there would be a problem based on the economic status of those who were offered a program in the first thing if they were not very, very similar. If I am marginally wealthy and I'm offered a 35% discount, I'm more likely to take out than someone who is broke and is offered a 35% discount. Then that would affect your sample.

PROFESSOR:

With one twist. So it's not the levels that matter, but it's actually the slope that would have to matter. It has to be not that the wealthy is more likely to take up with any given level. It's that they have to be more elastic or less elastic than the poor in order for that to be an issue. And then you're absolutely right, that is an issue. And then what you're studying is when you do that subsidy, you're studying your treatment effect on those people who are going to be more responsive to that subsidy. Is there another hand? No. Wendy, now we're to your question, multiple treatments.

So let me just say one more thing on encouragement designs. So one of the key things to remember with encouragement designs is that in a lot of situations, the encouragement design is-- in some situations, it is set up where the control group does get into a program. So where you're dealing with a 10% take up rate in control, and a 30% take up rate in the treatment, In a lot of the setting, though, it's really more that you just have incomplete take up in the treatment group, that participation is voluntary. And so by encouragement, all we really mean here is that a treatment is being offered to people. They can say yes or no, and they're not being offered to the control group. And so we end up with take up of some percent in the treatment group, and zero in control.

Like the savings experiment that I just referred to a moment ago, we had a 28% take up in the treatment group, we can't make people open a savings account. All we can do is offer it to them, and we had a 0% take up rate in the control. We did a similar thing in the same place in the Philippines on a commitment account to stop smoking. Again, 11% take up rate in the treatment group. We can't make people want to stop smoking and sign accounts and contracts to do this. But we can prevent the control group. So there was perfect control in the control, in the sense that they were not offered the opportunity to open the account.

But the treatment group has to be voluntary. That is what it is. And so it's an encouragement design, with 11% percent take up rate in the treatment and 0% in the control. So sometimes you do have control on one half but not the other for who uses. Yeah?

AUDIENCE:

And so just the main point is that what constraint are you going around by an encouragement design? Just an ethical problem if you can't afford treatment to everybody?

PROFESSOR:

In that situation, I don't know that I'd pose that as an ethical issue. But the point to be made is just that you can't force people into a program. It's a voluntary participation, and that's OK. So one of the things that I've often read or heard is when someone says, well, wait a second. This is voluntary participation, so how can you-- doesn't that introduce selection bias? And the answer is no, because what we're going to do when we do the analysis of that, is we're going to compare everybody who was offered the account, everybody who was not offered the account. And so there's no selection bias there.

There would be a selection bias if what we did is we analyzed everybody who took up in the treatment group, and compared them to everybody in control. And that would be a flawed analysis. But that's not what we do. So if you ever hear someone say, ah, encouragement design, doesn't that introduce a selection bias, because participation is voluntary? The answer is no. That only introduces selection bias if you do the analysis wrong. What you want to do is compare what's referred to as the intent to treat analysis, and it means comparing everybody who's offered to everybody who's not offered.

Multiple treatments. So this goes back to Wendy's question. This is one of the areas where I tend to think is most ripe for helping-- going back to one of the first points I was making about making sure that the evaluation speaks nicely and informatively to needs of the implementers, needs of the organization. That a lot of times, there's very specific operational questions that they have. Should we really do it this way? Or should we do it this way?

I really made some tough choices here, and I just went with what I thought was best. But gosh, if the research can actually help guide me and tell me whether this particular component is necessary or not, that would be great. So imagine you're doing-- let's go with Wendy's example of an agricultural program. And suppose that you're trying to decide, how important is this training component? I'm going to provide seeds, and introduce people to marketplaces. I'm just making something up. Let's not get into the details too much, but let's just say there's a training component alongside of it. And that training component is really expensive, it takes a

lot of time.

And I'm thinking to myself, OK, I can help twice as many people and drop the training, or keep it my current program size and have training. What's better? Well, the research can help answer that question by having an evaluation which evaluates the overall program, but then also randomizes whether or not there's training involved.

And so this is one of those key areas where it's a win win for operations. Where you can help answer questions for them beyond the simple impact question. There are situations in multiple treatments that we've been in where there's no pure control. And there's nothing invalid about doing that. It does validate the study, but we just have to remember that you're no longer saying, what is the impact of the program compared to not doing the program? You're now comparing it one option relative to another option relative to another option. And hopefully in the design there, you have one option which is kind of like a placebo, so that you have some group that you really don't-- it was in a very extensive way, but you have some method of being able to say what the overall effect is.

But there's many situations we're in where that's actually just not part of what the study's about. So we've done savings product designs are a perfect example of this, where we're dealing with people who opened a savings account. There's no control group of people who were not offered a savings account. We're just dealing with a bank and they take people in. And the question to us was, well, how can you help our existing savings people save more? So we tested something out in three different countries where we sent people reminders to save. So half the people basically got a little text message saying, hey, don't forget to save this month, and half did not.

So we have no control group here of people who got no savings account. So we're not measuring the impact of savings on things. We're just measuring the impact of getting this reminder on how much you save. And we've done similar things with loan repayments. There's no study on the impact of the credit, we're just testing out operational questions about how to run the program better. And in those types of designs, we'll often test out five different messages all at the same time.

I think I said that slide. Oh, maybe not. Yeah, we talked about randomization in the bubble.

So this is the list of the various things that we've now described. And just remember that these are not mutually exclusive. Multiple treatments and encouragement design in particular kind of

fit within almost any of these other things going on here. Any questions so far?

OK, part two, gathering support. So here are some things that we commonly hear. So this part of the lecture is really all about how we deal with this kind of introductory conversations, exploratory conversations where we're trying to work with partners to figure out how to go about doing a randomized trial. So one answer which is always a tough one to get, but I already know the answer. And I don't want to risk learning that we do not have an impact.

There are situations that we'll be in-- and I don't mean to sound like a pessimist-- but there are situations we'll be in where you just realize this is not a good setting for it. You have to work with people who actually want to know the answer. And you can recognize that merely observing that their program has grown is not necessarily a sufficient measure to say whether they've had an impact. And it's certainly not a sufficient measure to say whether their program is a good allocation of resources compared to other programs that have also had similar operational success. And so when we have to make the tough choices, this is where we need the evidence.

Listening is probably the single most trite but important thing I have to say on how to have these types of conversations. Trying to understand the perspectives and the objectives of the people in the table. What is it that's making them tick? What is it that's making them have this conversation in the first place? And finding ways to make the research operationally useful is perhaps the single most useful and important thing to do when working in the field.

One thing that I've often found too is that often in practice, there's a caution. There's almost a mistrust that some might have if they're not familiar with what is that's going on. And it's one of the most important things that the field staff can do in working with the organization is to just gain the trust of the people in the field who are working for that organization. And some of this comes about in getting their feedback and input into things like survey design. Having it so they feel part of the process, and their input is received and incorporated into what we're doing.

And that's good for the program, good for the evaluation to get their feedback. It's also good in a purely interpersonal way, in terms of helping to have a relationship that's good by making sure that people feel that they are part of that process.

So some other specific things that come up. The first, one of the most common things is gossip. People will talk. So what do we do if the control group finds out about the program? So

I think the thing to think about is to try to separate out these types of issues. Let's just put this into a more general category called spillovers. So spillovers meaning there's any sort of indirect effects that are going to occur, from those who were treated to those who are untreated. I think it's really important to separate these into two categories. There's natural spillovers, and let's call them research spillovers.

Now, by a natural spillover, what do I mean here? I mean a spillover that is naturally occurring. That is, if you go and you provide a service to 100 people, the fact is this is going to affect those 100 people and 200 more. And that's the nature of the intervention. It has nothing to do with the research. The example that you have a case on is deworming. We're going to deworm half of you. The other half will benefit from that, because you're going to be less likely to catch the worms from the first half.

Let's say I took half of you right now, and I went into the other room, and I gave you a whole big lesson in power calculations, and I ignored the other half, and I didn't give that to you. Well, there'd be a spillover. You'd come back, you're in a group, you'd talk. Oh no, no, I just learned about power calculations, let me show you. Hopefully it'd be a positive one. So these are all natural spillovers though. There's learning that takes place. You teach some people, they teach others. You deworm schoolchildren, other schoolchildren benefit because they're less likely to catch the worms.

There could be negative spillovers. We go and we offered really, really cheap credit to some people, or we only offer it to some because we're constrained. That's the organization. That's just how many loans we make. It gives them a competitive advantage-- I'm not saying this is right. But this is an argument that people will make when they are arguing against subsidized microcredit. And what does this do to the people who don't get access to the microcredit loans? It shuts them out of business. It makes it so they can't operate their enterprise because they're competing against someone who's getting subsidized credit. So that has a negative spillover. These are natural, though.

So a good study is one that helps to measure these things. And there are ways that we can design experiments to measure those types of spillovers. So a very simple example of one that measures this is suppose we have villages. And what we're going to do is take 90 villages, and instead of just dividing them up treatment, control, what we're going to do is we're going to divide them into three piles. We're going to divide them first into two, treatment and control. So

we'll have 60 of those villages being treatment and 30 being control. And then within the 60 that are treatment, we're only going to go and deliver services to half the people in those villages.

So what do we have? We have a treatment village that's half treated, half untreated, and we have control villages. Now throw away the people that got treated. Just ignore them. What's an interesting analysis to do here is to compare the untreated children or people, or whatever the intervention is in the treatment villages, and compare them to the control. These are two people that didn't get services. Neither one got treated. But some of them live near people who got treated and some of them do not. So that measures the indirect effect. That measures the natural spillover.

So if a natural spillover is something that one was concerned with, this is exactly the way you would think ahead of time about setting up the research design to measure that. But then there's unnatural spillover, what I was referring to as research spillovers. Yeah?

AUDIENCE: Just a quick question. Does the fact that now the treatment is half the size compared to the entire control group, does it matter?

PROFESSOR: Yes, it matters. And that's a question of power calculations. And so you have to trade off your measurement of the spillover versus your measurement of the direct effect. But that's a mathematical problem that can be solved analytically.

So research spillovers, those are the bad ones. These are the ones we don't like, because it's not interesting, it's not useful. It's not representative of what happens in the real world when you do an intervention. It's just an artifact of the research process. The simplest example is the control group person who says, I don't like the fact that I am in the control group. Maybe they don't believe it was truly random, or they just don't like the fact that they didn't win the lottery. And so they actually change their behavior now because of this.

Let's use a very simple example of being in a bank, and let's say you're doing a lottery across existing borrowers, and half of them got an extra service to go along with their loan, and others did not. The ones who didn't get the extra service, they're now upset. I didn't get the extra service, I'm not happy. Why did they get it? I didn't get it. And you can explain, well, it was random. They don't accept that that. And now what do they do? Maybe they don't pay back their loan. Maybe they leave the program altogether because they're mad.

Now in studies we've had, I can honestly tell you we've not had this happen yet in a microcredit setting, but there's certainly things that we will do to try to avoid the problem. So for instance, one of the studies we had where this was a bigger concern than others was we were testing out group versus individual liability. Now most borrowers really like the idea of individual liability if they're given a choice. They don't want to be on the hook with other people in their community, they much prefer to have a loan that's just to them and them alone.

So when we were randomizing whether people got offered group or individual liability-- and it was existing borrowers who were already borrowing from a bank-- what we had to do is take villages that were really right next to each other and put them together. Because we couldn't have it that we had these little sister villages where there was lots of interaction across, and one got switched and the other did not. So we put them together and treated them like one.

And so basically, it's another way of saying this is when you think this is an issue, you just need to think about making sure that you have some sort of boundaries separating out your treatment and your control areas. Now in an urban setting, it could be a little bit harder if you don't have clearly defined boundaries. But it's still feasible to do this type of process. You just have to think a little bit about how to do the boundary, and also what to do if you have control group people who do come in. And so this is an area where encouragement designs might actually be a useful way of dealing with it, is to allow some control group people in, for instance, if they come in. But otherwise not unless they actually come on their own.

So this is another way of saying take an urban area, and you just encourage some blocks to borrow and not others. Encourage some blocks to go to school and get some extra service, and not others, whatever the program is. And that's a way of trying to make sure that if the groups talk to each other, it's OK. It's not going to ruin the study, and there's no jealousy, it's just a matter of some receiving encouragement and others not.

So fairness, hopefully this is-- the one point, if I had to leave you with one simple thought of this lecture, it's the fairness point. It's perhaps the single most commonly raised issue, and it's the easiest of the issues to explain in 99.9% of the settings we're in. And it's this fairness issue of, oh, but gosh, I don't want to do it by lottery, I want to do it by some other process. And the answer-- or gosh, I can't imagine restricting access to people. And the answer is always just about the same, which is how many people can you deliver in this program? What's your budget? Now let's divide by the cost per person. And so you can do this for 1,000 people, or 2,000, 10,000, whatever your constraint is, you have a constraint.

Now all we're going to do is use that constraint to then find a way to do this randomization, and that's it. So we're not restricting access to anyone. The only sense in which we're restricting access is perhaps a bigger picture or thought, which is if half a million dollars is being spent on the evaluation, that's half a million dollars that's not being spent on services. That's the only sense in which the randomization is costly in terms of delivering services. But this is a different calculation all together.

This is now asking the question of whether it's worth half a million dollars to find out the impact of this program or not. And that's just a very different question. It's not about the resources and the fairness to those individuals, it's about the question of whether this program will be done enough times in the future, and will the marginal value in terms of our knowledge about what's being learned from this study be high enough to warrant spending the money on the research period. It's a different question.

So on ethics, some of the things that we'll often hear are statements like, well, it's wrong to use people as guinea pigs. Or if it works, then it's wrong not to treat everyone. So the first thing to think about is that first of all, it's not-- I think it's a very leading question first of all. One thing to often ask yourself-- or ask people in this type of conversation-- is why is this different than prescription drugs? Why should we be more willing to proceed and deliver interventions and deliver services to people without knowing their impact then we are to prescribe drugs?

We ourselves, for instance, would never take a prescription drug if it hadn't gone through a randomized trial, or more than one. And so why should we be using a different set of standards in terms of the ethics of the two, in terms of the bar, the rigor that we want to use in order to decide how to allocate our resources.

The second thing to think about in terms of, if it works, then it's wrong not to treat everyone, I think there's an important point to note, that there's lots of ideas that-- there's two issues that often come up in this setting. First of all, there's lots of ideas that sound good, but then when evaluated, turn out to not work. And even when something works, the question is how well does it work? We have other ideas that work. So even if something is good, even if everyone around the table is totally confident that it's going to work in some respect, we don't know how well it's going to work.

And when we're allocating resources, we're not just trying to beat zero, although that's always good. We're actually trying to do the best we can. And so we're choosing across five ideas,

and they all sound good. No one's throwing out ideas that-- well, I shouldn't say that, that's probably not true. We can probably think of some that don't sound good. But for the most part, we like to think that we're sitting around the table thinking across choices that sound good, and we have to choose. And so that's the most important thing to remember in that type of conversation.

Cost. So there is two things that often come up when people talk about cost. And there is a common perception and argument that randomized trials are much more expensive than other approaches. So I think there's two things to think about when this type of conversation or point is made. The first is thinking about what the cost-- it's about doing a cost benefit analysis. So let's assume for a second randomized trials are more expensive. And I'll point out some examples in a moment where it's not. But let's say it is in a given setting. Well, that's only part of the equation. You have to say, well, what's the cost and the benefit? The whole point of doing randomized trial is to think about costs and benefits too.

There's no reason why we should think any differently when we think about how to evaluate. So what's the benefit we're going to get them from doing a randomized trial versus a non-randomized trial, and what's the cost difference? And if the benefit that we're going to get in terms of the reliability in results is high enough to then make more impact on our ability to make future decisions, well then, it's probably worth spending a little bit more money on the evaluation itself.

Now obviously, that's relative to our existing knowledge in the space. If something's already been tested 15, 20 times, then this might become a situation in which you would argue that no, the benefits don't outweigh it, because the marginal impact from the research on one more study is not going to be that high. And so the costs are not worth it. I'm not aware of situations I would say fit that, but hopefully we will be there someday.

The second is that the cost of doing randomized trials is often not actually more expensive than non-randomized methods. But it's really key to state clearly what the counter-factual is here. What's the alternative method one's describing? So it's clearly cheaper than doing nothing. And there are situations that I've been in where my best advice is don't evaluate. For whatever reason, the setting is-- you're not going to get a reliable result, and the best thing one can do is to not do the evaluation.

Let's compare now to the most common comparison one makes, which is to a non-

experimental quantitative method. So suppose the alternative approach is to survey a bunch of people who received a service, and to survey a bunch of people who didn't receive a service. It wasn't done randomly. It was some people chose to be borrowers from a microcredit program. For those of you who know me know I do a lot of work in microcredit, this is why I keep using microcredit examples.

So I'm going to survey a whole bunch of people in microcredit that are part of a program. And then I'm going to go into the same community so I can find people who seem very similar, have the same macroeconomic conditions, but are not participating in the program and going to survey them. I'm going to follow everybody before and after.

So this is actually a more expensive study. Why? Well, I need a larger sample size here. I need a larger sample size because I actually have to really understand something deeper now about who's opting in and who's not. And I have to try to use my econometric tools to correct for selection biases. And this costs me sample size. And so this study would actually cost more money, because I would want a larger number of observations in the analysis in order to try to get it right.

Now it's clearly going to be more expensive. Now let's flip to another one. If our counter-factual approach is to not even do any surveys, like there's no big econometrics, but instead to do a simple before after. I'm just going to compare before after. So you've studied yesterday, you talked about before after, you went through some of the issues that you have that you don't know what else is changing the environment. But if that's the comparison you're going to be using, well then yeah, this is the more expensive, because you've got to survey control group people too.

That's an example where it's hard to come up with settings in which there aren't outside factors-- economic, social, environmental, health-- that cause changes over time in outcomes for people, such that a simple before after analysis is in anyway informative at all about the impact of a program.

Timing, OK. I'm going to try to wrap up quickly. The one thing to say about timing is it's very common to have a constraint where the organization is like, but we need the answers now. Randomized trials are no different than non-randomized trials that follow people before after, but they're certainly going to be a lot longer than things that simply look retrospectively at people that have already received services and hold focus groups and discussions to try to

assess impact. And there's no way around that. So this is a question of just being patient, and working with organizations that are able to be patient in order to have those answers.

I'm just going to run through this initial slide so you can have the basic key points of the overall plan when we're doing an evaluation. So the three steps we have listed here are plan, pilot, and implement. I think it is important to note that there are situations where we don't actually do a pilot. Depending on the circumstances, the situations in which we do pilots are typically when there's a lot of uncertainty about what the intervention is in the first place, and so we're actually working with the organization to figure that out. Or if there's some uncertainty about the way the process is going to play out.

Maybe we're uncertain about the encouragement design. We're not sure if it's going to work. We're not really sure, will this actually encourage people to come in and use a service more so than they would otherwise? So we need a pilot to test out whether that approach will have an effect or not. We just need a smaller sample just to gauge whether we're dealing with 60% take up rate in our treatment group as a result of encouragement and 10 in the control, or are we dealing with 12 and 10. What's our range?

So the five steps we've laid out here is identify the problem and proposed solution. So I think one of the things that should never escape us is that you don't take a randomized trial. You don't start off saying we have a tool, now what research questions can we ask? You go the other way around. You want to think, well, what's the research question we're asking here? What is the problem that we see in the market we're in, in the society we're in? What's the market failure that this intervention is trying to solve or measure or test, and then what's the proposed solution? Think totally abstractly, don't even get into what the randomized trial is and how it will be designed. Just think first order about what the market failure is, and what the logic is behind the proposed solution.

Second, and this goes back somewhat when we're talking about in terms of identifying the key players. There's nothing more frustrating than a really good project, where you just don't have the right players on board participating and collaborating in a cooperative way to make the project work. Identify the key operations questions to include in the study. This goes back to hopefully the other theme of my lecture this morning, is about making the research into win win opportunities, finding those operation questions and trying to build them into the research as well.

Then design the randomization strategy, and define the data collection plan. Data collection can be done continuously lots of waves, one wave at the end. There's lots of other tools that we can use in data collection, both qualitative and quantitative approaches. One of the other common misperceptions that I've heard is people saying that there's a spectrum between qualitative and randomized trials. That's kind of mixing apples and oranges.

Qualitative versus quantitative is about how you go about measuring things and what you measure. A randomized trial is just about identification of the effect of an intervention. It's about random assignments of treatment and control, but it has nothing to do with whether the measurement is going to be done through a qualitative or quantitative process. And there's a lot of examples of studies that we have that use mixed methods and creative approaches for measuring things, and there's a lot of studies we have where it's very cut and dry, normal quantitative, how many potatoes did you eat type questions.

So pilots vary in size and rigor. The pilots and the qualitative steps that often go into them are very important for helping to understand the intervention and design it, particularly when we get into designs that are doing sub-treatments. A lot of times those come out of the qualitative process in the design of a study.

And then for the actual implementation-- oh, I skipped something. Identifying the actual target population is going to be covered later in the day, in the second lecture. And then collecting the baseline data will be discussed later on. When we do it, we usually do it, but not always. The actual randomization, there is various times and points in which you do it. This is what we were talking about in the beginning of the class, real time randomization like the credit scoring all at once, villages known up front, and you randomize them in or not to a program.

Then the next phase is implementation intervention to the treatment groups, and this is where internal controls can be really critical. There's nothing worse than doing all of this work, doing all these surveys, and then not having the right control in the field to be working with the individuals from the organizations that are delivering services to make sure that things happen the way they're actually supposed to happen.

And I've had projects go bust, where we're working with organizations that thought they had the right internal controls in place. And when we go in to do spot checks to see, and we go to some villages to see, are they getting services or not? And lo and behold, they were not. Or they were when they shouldn't be. And we go back and we try to work with them. I've had at

least one project I can point to that literally we just canceled after a year and a half. It was very unfortunate, but this is what happens when there wasn't the right level of internal controls in place. And I learned.

And then measuring the questions. One of the most common questions we get with measuring is, how long should we wait? And there's really no one answer to this. There's often a trade off with operations. If there is any sort of holding back of a control area, then this is going to be something that has to be negotiated and discussed with operations. In a lot of situations we're in though, it's not that the two sides are actually differing-- I mean, the operations maybe-- but the head of the organization might have incentives that are perfectly aligned with the researchers. They want to wait long enough in order to make sure that they've given their program a full chance to have its impact.

And so usually when I am posed with this question by an organization, I usually just ask right back to them, well, you tell me. What do you think you need in order to see the impact of your program? If you're telling me a story about it being a 5, 10 year program in order to see everything flourish, well then that's your answer. If you're telling me that this is like an amazing thing that just transforms people's lives within six months, well then we can go in six months and see that amazing transformation.

We might also want to see the two year impacts, but that would be something that could happen with the organization, and they could say, yes, we think it's transformation in six months. And two years is just beyond-- I mean, I don't know what the word is to say beyond transferring.

Lastly, analyse and assess results. And obviously, there's a lot more in the class that will be discussing that.