Rebecca Wolfe:	Good afternoon everyone. Thanks for coming. So today we're gonna have a discussion about how we can bring more rigorous methods, evaluation methods, to maybe some of the softer side. I think a lot of the presentations here are about health or economics, but I think this conversation will be about how can we bring these methods to kinda not softer, but maybe harder to measure outcomes in the realm of governance and politics.
[00:00:30]	
	I'm Rebecca Wolfe. I work for Mercy Corps on our research and learning team, but I also am a member of EGAP, the Evidence for Governance and Politics Group, which does field experiments looking at these issues and have run a number of initiatives on how to do this. I'll be moderating this session but with the rest of our panelists. He's Chris Fariss from University of Michigan. He'll talk about a project there. Morgan Holmes, who is the evaluation specialist with the DRG Center. Erik Wibbels from Duke University, and James from Georgetown University.
[00:01:00]	
[00:01:30]	So, everyone is gonna talk a bit about their projects. Morgan will talk about the strategy of the DRG Center, and then we'll all foster a discussion.
	So, Morgan, you wanna start? Yeah.
Morgan Holmes:	Thanks, yeah, and I'll just take a minute from my time to mention that Rebecca's done amazing work with Mercy Corps over the last I don't know how many years since I've been doing this, to really make them a leader in rigorously evaluating their work, which is like she mentioned what some people might have considered softer, fuzzier in some of the conflict mitigation area, and they've really been a leader in doing that of their own volition so I appreciate that immensely. But, yeah, the DRG Center, it's the Center of Excellence on Democracy, Human Rights and Governance, and it's sort of had a Renaissance from just being USAID's, just being USAID's Office of Democracy and Governance, to a center of excellence about six years ago, and one of the major changes that was made at that time that made it more of a center of excellence was dedication to generating knowledge and learning about this area, which at that time had, I can think we can fairly say, fewer randomized control trials or quasi-experimental evaluations of our sectoral programming.
[00:02:00]	
[00:02:30]	
[00:03:00]	In the last six-ish years since ... I think these are ... We are not causal of the other, but this happened in parallel with the academic world, that in the last six years or so, the amount of RCTs and quasi-experimental designs of the Democracy, Human Rights and Governance sector has risen, literally exponentially, from maybe in the 20s to more than 300. We did a evidence gap map with 3ie. I don't know if anybody went to the 3ie session. There are print-outs of it floating around sort of to look at the landscape of all of the impact evaluations done in the sector, and they came up with at least 304 of really excellent quality so I know there are at least that many now. So it's really a burgeoning field and we are, in our way at USAID, trying to contribute to this by embedding more RCTs and if necessary, quasi-experimental designs into our programming, and we have a unique way of doing this that we figured out over time. So, I just want to describe it for you real quick before we
[00:03:30]	
[00:04:00]	

pass it on to the actual principal investigators of the research.

| | So we've learned that if you try to contract, award, begin an evaluation at the same time as the program you're trying to evaluate, it's not gonna work, if you want to do a serious job of building in any kind of randomization or any kind of sort of more rigorous designs. So we start working with the USAID Missions who are interested in doing this sort of learning project two, two and a half years, before the program that we're interested in evaluating. It's a little bit of a misleading term, but they were interested in using for research, is even awarded. So we start way back and we bring these Missions together with academics, like James and Erik and Chris and their colleges who work on these subjects in academia. We find one of these academics who has a really strong publication record in the field and sit them down for a week long, intense session with the USAID Missions, and each afternoon of that week, we break out, and they hash out a design of what would ... First of all, what are your main research questions? What is your theory of change? Is that theory of change really practical? Are you really gonna see effects. They talk through all of this design with the USAID Mission officers, in that in and off its own right is worth the clinic, which is what we call this week, right there. |

[00:04:30]

[00:05:00]

[00:05:30]

| | But then, also during the week, and this is like the selfish part on our side, is we build in an experimental design into that program. So as it's usually part of the program, not the whole entire thing, but for instance I saw Niels from our USAID Uganda Mission walk in, so I'll use that example as part of a very large democracy in governance program there. We just looked at one particular mechanism of using SMS messages to directly contact your local district authority about public service delivery problems in your area. So we would just test a question related usually to one component. So we use this clinic where people come together, they learn about the academic evidence in their field. They use that to develop the program design. They think hard with the academics about their theory of change and its realism, and then also build in some experimental design into the program. So these guys are gonna talk about the interesting projects that they have done in our regard. |

[00:06:00]

[00:06:30]

[00:07:00]
Erik Wibbels: | I think I was the canary in the coal mine, in a sense because it was right about 2011 that the DRG Center contacted me and three other academics with the idea being that we need these outcome lead experts in different pieces of the governance portfolio. With the idea being that the academics were gonna help figure out how to make this kind of program and evaluation more serious and more rigorous, and we flailed for a while. We started trying to work with evaluators on the quality of the evaluations, but that seemed like not a great idea if the underlying program wasn't so great. We tried lots of things and we settled on what ultimately became the model the Morgan just described, and this project might be, the one that I'm kinda describing, Ghana might be the first one to emerge out of that model where this is basically me and my Co-PI sitting in the mission with the DG officer and figuring out what they were interested in and how could we turn that into a rigorous design, and so, here it is.

[00:07:30]

[00:08:00]

[00:08:30]

[00:09:00]

I'm gonna give you some preliminary results that the endline has just in and these are not publishable quality, they're preliminary. Okay, so this is the basic problem that the project is aimed at solving and that's the districts in Ghana spends, something like 3% of GDP every year building stuff, and the stuff that they build, if it's built well, is incredibly important for well-being. Things like six-room schoolhouses, that's this sort of thing, health clinics, public toilets, these sort of things that can have a huge impact on well-being. One of the inefficiencies is that something like 35% of projects are never completed every year, and so that amounts to a waste equivalent to 667 additional three-room schools every year, and this is not the only form of waste. If you dig into the project budgets you see that the ... So this is a spec schoolhouse, and the cost of the spec schoolhouse varies by a factor of seven. There's a sense that something was amiss in the delivery of these development projects that if done properly could be incredibly impactful on lives, and this was combined with a sense among citizens that district governance wasn't working for them. And so the project aims at these things.

[00:09:30]

[00:10:00]

So the design here is we have a 150 districts, we randomly assign them into two, treatments and control. This was designed as part of the clinic sort of model that Morgan just described. It's not unique, but one of the things that's distinctive about it is its government to government. It's been extremely interactive between the academic side, the evaluation side, the mission side, the implementers' side and things have gone wrong sometimes as they always do and because of the interaction, it's been possible to catch those things along the way and now we're at the point of evaluation.

[00:10:30]

[00:11:00]

What are the treatments here or what's the programming? We have essentially two forms of auditing. One is a top-down audit, a performance audit, this isn't a financial audit, this is going in to "How did this project get into the budget? How is thing contracted? What is the quality of the actual construction and what's the NUs like?" So all the way from there, beginning to the end, and this was done via training a new team of auditors in the Ghana Audit service. So the Ghana Audit service had performance auditors but they were all focused at the national level, this is a team that was built specifically to do performance audits for districts. The results were then turned into scorecards and disseminated in assorted ways and we're interested in that sort of very top down kind of audit versus a social accountability audit where we, Care International and Partners, we're training citizens on the ground we're these projects were being built and how to monitor them and what to when things weren't going well. And the sort of thing that might not go well is if take a typical Ghanaian brick and you drop it from knee height, it can break into three or four pieces and that's a good brick. If it breaks into a whole bunch of pieces, that means the contractor is skimping on concrete.

[00:11:30]

So things like that, that are really simple but really important and then providing communities with the means of for instance, "This is the phone number that district engineer." So the results of these social audits were also turned into scorecards and have been disseminated, and we have initial findings. So there are three target populations here really. There are citizens who we hope are gonna become more involved and generate more accountability on these projects. There

[00:12:00]	are the district politicians, the district officials, who make budgetary decisions on what projects to build where, and then their administrators who once that decision is made, they're responsible for all the contracting and making sure that these things are actually implemented. And so they're different pieces of this project that aim at this different populations. I'm gonna show you just some results from the administrative endline survey.
[00:12:30]	So the first and I think the most important finding here is that ... I'll just prime us by saying, these effects are actually pretty large. I'm not gonna give you the nulls, there are some nulls on some things, but a really important thing is that there are these internal audit committees that all districts are supposed to have but many districts don't have. And one of the things that we have seen is over the life of the project, we're much more likely to see a functional and active internal audit committee. That's what this Erik thing is here. For me, this is really important for sustainability purposes because this is something that is internal to the district, the machinery itself, and this is not built by the project. This is something that the districts build themselves, into their district administration.
[00:13:00] [00:13:30]	The bottom-up audit, the social audit, is generating some pretty substantial effects in terms of the perception of administrators on how engaged citizens are and it's also having a pretty big impact on how they spend their time, so two plus hours a week more on responding to citizens as it bears on these kinds of projects. Then we get two really interesting findings where people are sort of getting pissed off as a result of the results, but this is actually good from the point of view of accountability. So when you get these bottom-up audits, there's a perception that the current projects are deficient on assorted measures of project quality. That's when you know that project quality is bad, you can actually do something about project quality being bad. This is useful.
	The top-down audit, which is to say the Ghana Audit Service audit, we get the same thing, but we have districts now comparing themselves to each other, so there's a sense that my district, because of my audit, I get this sense that my district is not doing as well as other districts. Again, this is potentially a mechanism for improving outcomes.
[00:14:00]	Okay, so now we have an extension, that I'm gonna talk about quickly. Am I okay on time?
Rebecca Wolfe:	Probably maybe another minute or two.
Erik Wibbels:	Okay, so it could be that this project also has implications for where projects are placed by district officials, in the sense that when you are in a non-accountable environment, it could be that you put projects in places where your supporters are as opposed to where people need them. If all of a sudden citizens are more engaged and informed on these matters, it could be that you put them in the places that need them. So on top of this, we have built this crazy project, which is ... So this is a very small bit of our data, and we have geocoded ... So one of the nice things in Ghana is that the districts have to prepare a annual progress report that
[00:14:30]	

[00:15:00] lists projects and cost and where they are so that they can get the following year's transfers from the central government. We've digitized more than 45,000 projects now and have combined this with voting booth level data. Voting booth level data, not district level data, and enumeration area level census data. We basically have the census, so at a very nuanced level, do we have measures of need and measures of sort of the political characteristics of places, and we can do all sorts of cool things.

[00:15:30] This has become now a partnership with the Ghanaian government where this is the foundations for what is becoming an infrastructure map for the whole country. So these are, whatever. It's a fancy looking map that I don't have time to explain. One of the things that we're discovering is that there are lots of problems in the base maps that the government uses, and so this is being used as a process for fixing that, expanding the geospatial capabilities of the Ghanaian government. And one of the cool things here is, all of this sort of important stuff that's going on in the government is also generating I think ... I don't know. Maybe you would disagree with me, I think it's gonna be the best academic data set, the biggest at

[00:16:00] least, on sort of the politics of distribution in space. Why do politicians put stuff into places that they put stuff? I think answering that question is important, and I had other stuff that I was gonna say but ... There's more, because now we're working with the Mission and the Ghana Audit Service on potentially designing another RCT with regards to the expansion of the audits in new and interesting ways.

[00:16:30]
James H.: So I wanna start with I guess the question that motivates the work that I'm doing in Uganda with Stuti Khemani of the World Bank and Claudio Ferraz at PUC-Rio, which is the level of public goods that countries like Uganda and others like it provide seem to be very low and the quality is very poor. And one symptom of a poor public goods provision is high levels of [inaudible 00:16:53].

[00:17:00] Now I think we started, and in some ways I'll say that Morgan's been very influential in that she's getting us into the field. I think this work started with the report that Stuti, Claudio had written around essentially kind of a theory that in fact one of the reasons bureaucrats in these environments perform very poorly is in fact that they're structured, the incentives and expectations are structured by what's going on in the political market. I guess in most of these environments, public

[00:17:30] goods provision is managed in this way. The bureaucrats do the work. Politicians, I guess, draw the budget and monitor what the bureaucrats do. And so if the politicians incentives are in fact to target, very much in the sense that Erik is referring to, to target certain private goods perhaps, maybe not even public goods, as opposed to public goods then of course the bureaucrats may in fact sort of infer that their performance is likely not really important and therefore kinda the

[00:18:00] monitoring level are going to be very low.

So that in some ways is kind of the broad motivation for the work we're doing, which is to say their political norms can affect professional norms. And I think one of the things that we're very excited about and hope to be able to do is actually try

[00:18:30]	and measure professional norms. Political norms, ultimately, I think down the road, we'd like to work on that as well. But I think one of the things that's really important about ... You start with some high level theory and how do you translate this onto kinda research on the ground, is that the actors who're actually delivering programs provide a really powerful platform for researchers like us to actually sort of measure these things.
[00:19:00] [00:19:30]	And so, there is a great program in Uganda called the, I don't know, the GAP ... Governance and Accountability Partnership program, I'm sorry. And this program gets politicians and bureaucrats to sit down and actually talk through a series of ordered reports that are produced by the government. And in some ways, not quite the same I guess level of intensity that Erik is talking about, you actually get an opportunity for citizens to question why things are the way they are, and for the audit reports to in fact provide information to citizens and for bureaucrats in a very public setting. There's radio. And in fact, so when heard about this opportunity to build consensus, we thought, "This is what we want to evaluate." So we rush on and talked to RTI and to the USAID Mission there and say, "You know, you have a great program." Morgan of course is the one who made all of this possible, but RTI said, "Look, we have a particular structure. We work in particular districts. You guys want a particular research design. It's just not possible to do the evaluation you want to do."
[00:20:00] [00:20:30]	And so we actually went back and decided, "Okay, maybe we can actually change the level at which we actually operate." So go from districts, there's only about 112 in Uganda, GAP only works in 35, they were only gonna be doing their program in 17 particular districts that they have already chosen so there's no scope for kinda a lot of the usual I guess measurement strategies that we like. But they were very open to the idea of saying, "Well, you know, if you really want to measure norms and if you think that in fact there's a way in which you can actually design a program to change people's expectations and teach us professional norms, you know, why don't you do it at the smaller levels. Do it at the school level, right? So let's see if we can actually get teachers together, deliver a deliberation program to actually sort of change both social and empirical expectations and also use some signals from district leaders and politicians and see if in fact, you know, you actually produce the changes that you expect to see based on your theory."
[00:21:00]	And so we're at a stage, I don't have any results. We are hopefully gonna launch a baseline survey in a month or so, and so I certainly hope to have an opportunity to come back and talk to you about it.
Rebecca Wolfe:	Thanks James. Chris?
Chris Fariss: [00:21:30]	Thanks for having me and thanks for inviting me to participate in this project, I can't remember, two or three years ago now. I wanna say a little bit about that teaching and learning clinic because I think that was really essential for the quality of the RCT that I helped design in Haiti. But first I wanted to talk about the context in the Haitian prisons. If we were at Michigan in the classroom I would make you do this but just imagine I'm making you do this. So luckily we're in a fairly confined space,

| [00:22:00] | but imagine that this half of the room stood up and walked over to that half of the room. And now imagine that you hang out with each other for the next year. That's about the size of the prison space that 60 to 70 people will inhabit in the penitentiary in Port-au-Prince. They're about the same. The number of prisoners in Port-au-Prince is much larger, but the circumstances for prisoners across jurisdictions in Haiti is about the same. |

Now imagine this has been six months to a year, you haven't been working, you're probably sick because of the conditions, and also imagine that you haven't actually gone and seen a judge yet. No progress had been made on the crime that you've been charged with. So the project that we're working on with Projustice in Haiti is doing many things. One of the things that it's doing is it's providing free legal representation to individuals in this prison, and that was the program that we were tasked with evaluating. I'll give you the happy ending, it works and for a very cost effective rate. Training and bringing in defense attorneys. Pushing these cases through the system actually works.

And so, there's a rule of law story here where on paper, the judicial institutions in Haiti should work pretty well. You shouldn't be in this prison space for more than 48 hours, on paper at least, before seeing a judge, and you shouldn't be in this space for more than 30 days before having your case adjudicated and the sentence decided. But as I said, in practice, this isn't what's happening, and so with Projustice we were able to develop this randomized control trial where because of limited resources on the Projustice side, they can only provide this service to a subset of the detainees that were eligible for it. We randomized the ordering in which the services were provided.

If you go into the, I think it's the ... Across the way, there's a booth with Nork, they have the impact evaluation because we're done with the impact evaluation, so you can read about it. And this paper, that's over in the market place, or it's online as well. So, once this impact evaluation started, the detainees received the legal service provision in the order that we randomly specified, and were able to show through this just very slight modification to the program that they wanted to implement in the first place, that there is a very effect of case advancement through the system. People started having their cases adjudicated. They started making their way in front of judges. Judges started making decisions about what they had done. In many cases, they had already served enough time, and so they were released. In many cases, too much time. But we showed that this process worked and we showed that it's cost effective.

This gets a little technical. We did a lot of simulations to assess the cost effectiveness of this program, and in the worst case scenario, the prisons were substituting away dollars to house prisoners that we could use to pay for their release. So in the worst case scenario, Haiti would be paying the same amount to pay for these legal defense teams to go through and help people get out of prison as they are to house the prisoners, and there's a lot of poor outcomes with respect with the individual's health and their economic opportunities when they're placed in custody here.

[00:26:00] [00:26:30]	I'd like to just take a second to say that, in my experience with the teaching and learning clinic. We wouldn't have been able to develop such a fantastic intervention, research design, without the help of the USAID team from Haiti and then the Projustice team. And we met and worked together for, it was actually quite exhausting, it was 40 hours over five days. I remember lots of coffee breaks just trying to keep the energy up because we were learning with each other in the mornings, and then we're sitting down in the break out sessions in the afternoon thinking about how we could design an impact evaluation that would demonstrate that this program was working. And it was really a fantastic opportunity and thanks again for inviting me to do it.
Rebecca Wolfe:	Morgan do you want to come up?
Morgan Holmes:	Sure.
Rebecca Wolfe: [00:27:00]	Morgan, I'm actually gonna ask you the first question. How do you decide which DRG programs to do these types of more robust impact evaluations? I'm sure DRG has many of them, and so ... But you can't do this for all of them.
Morgan Holmes: [00:27:30]	Yeah, that's a great question. I wish there were a more technocratic sort of algorithm that we use, it's actually more human centric. We reach out, we have a great team in the DRG center that stays in close touch with all of the Missions that are sort of regional teams, so we involve them and say, "Hey, which are the missions that are in program design and who have people who are really interested on research." And because, like Chris was mentioning, you need someone in the USAID Mission, they're all so busy and so overwhelmed, to take on also managing and being involved with an impact evaluation where you have to go the extra lengths with your ...
[00:28:00] [00:28:30]	You can't just pass it off to your implementing partner and say like, "Run the program." You have to tell them, "We can't run it here. We've got to run it there. It has to look like this. It has to look exactly like this in all these places but not these other places." And you have to say on top of them to make sure that they do that, so you really have to have a USAID Mission officer who is passionate about research and willing to do all this extra work, and myself and all of us who learn from it are hugely indebted to them for doing that. So it really depends on if they're in program design at that particular moment in time, we do these clinics about once a year, so if that's the part of the project cycle that they're in and if they're interested in research.
Rebecca Wolfe: [00:29:00]	Okay, thanks. To Erik, James and Chris, so these are large programs multifaceted, how do you see it lining up both with your research interest and would it be easier to just find a way to do a field experiment without all of the USAID bureaucracy around it? I don't know who wants to answer first.
James H.:	So, let me take a crack at this. So I think the challenge in some respects is that it's always better to have more control over the implementation and the measurement

[00:29:30] [00:30:00]	in general. So it's much better to do a field experiment if you can do it. I think for a lot of the governance related evaluations, the implementation is actually sort of a pretty large lift and requires I think a platform that, personally an academic, I haven't been able to generate. So I think it's actually valuable to have people go on the ground who have the relationships who are actually delivering a program to actually work with rather than do it yourself. I think they are small, on the margins I think there's some interesting things you can do. Information in general that you can sort of build that yourself. But if you really want to produce information about what's going on in a community, I think it's much better to work with people who actually have the experience doing that.

Erik Wibbels:
I mean, I would say I completely agree. I do think, thanks to Morgan and the DRG Center, a lot of this stuff is getting better and easier, and that's great. I do think that there's one frontier out there, which is strictly on the data collection side. So forget about project design and forget about project evaluation. There's so much data that's being collected in all of these projects, and as academics we have lots of

[00:30:30] cool strategies and techniques and things that we can bring to those data collection exercises that would be great for us and would not add any cost at all and we could learn a bunch even leaving aside the issue of program design and evaluation.

Rebecca Wolfe:
[00:31:00]
Just a quick follow-up on that. Putting in the surveys, but I think it goes back to this issue of Morgan said earlier about the control. So Morgan said earlier, I also work on the implementers side, and it's that having our teams being willing in a sense have something static for so long that's more of the hurdle than adding in a measurement type of question.

Erik Wibbels:
What's static in this case?

Rebecca Wolfe:
So often a program design, in order to have your between baseline and endline, has to be really static. And so there's also a movement towards adaptive

[00:31:30] management in especially more fragile content. So to me, it's not so much adding in some measurement issues, it's how to do a randomized control trial for a time where when program teams want to actually change things. And so, I don't know if you have any thoughts on that.

Erik Wibbels:
Well, what I meant to say is that forget about the RCT. I think there's lots of cases we're I'm just like, "Forget it." But if you could just be involved in the design of the

[00:32:00] instrument, so here's an example. I was working with some RTI people. We ended up being a little bit too late but [inaudible 00:32:07] had a girls education project, and I think it was nine very, very poor settings. Afghanistan, Somalia. And a lot of the reasons that they thought was dropping out were sensitive questions about how the household behaves with regards to girls versus boys and early marriage,

[00:32:30] and they got all sorts of surprising results at baseline suggesting that there were no issues on any of these thing. Like, "Well duh." Like people know what the right answer is. We have techniques for asking those sorts of questions that can get us closer to what's actually going on and it would not take that much work. It's not about the programming, it's about designing a piece of the survey instrument.

Rebecca Wolfe: Chris do you have a ...

Chris Fariss: I'd just like to add on to that. From my experience there's a lot of record keeping that's happening in the context of the implementation of these programs, and a lot of that record keeping is paper based. And I think that there's a lot of academics out there that would love to have access to some of this information that's being collected as the program is implemented. That has nothing to do with necessarily with the RCT component of the program, and I don't know what the mechanism

[00:33:30]   would be but making some of that information available to academics out there would be beneficial to additional insights, additional analyses, additional research questions that might arise that could then be considered in an RCT setup.

Rebecca Wolfe: The follow-up on that Chris, so I think one of the other considerations with these more rigorous evaluation designs is that the team who's implementing actually

[00:34:00]   often doesn't benefit, but you were hinting at ways through some of these other types of measurements and things we could do, the team actually benefits as they're implementing and learning. And so, do you have any thoughts on that with the Haiti program?

Chris Fariss: I think that, for example I was thinking of there were boxes of case files that we hand scanned. Took about two days to hand scan I think four or five boxes, but there's many, many more boxes of files that have been generated prior to the start

[00:34:30]   of our RCT, which would have been really beneficial had that data been available to us as we were starting to think about the actual RCT design, and that would have required some, not much additional infrastructure but a desire to make that information available as the program started for academics who might be thinking about how to build the intervention.

Rebecca Wolfe: [inaudible 00:34:49] James or Erik, anything to add to that or ...

James H.:   I guess let me just add one thing on the control issue, because I think that's really
[00:35:00]   the sticking point, which is to say an implementer, they have targets of what they're supposed to sort of complete and sometimes the research actually slows things down. That's attention that in some ways has to be negotiated ex-ante, and I think that the more you prepare, I guess, before the program to essentially kind of establish what it is you want to do, and align expectations about how long this is gonna take and what's gonna require. And perhaps even perhaps emphasize that

[00:35:30]   the benefits, not just, you know, " We're gonna write a research paper." But we're actually gonna something about whether this program is working in the way we think it's gonna work. I think that may actually help, but that attention is certainly something ... You know, I can't build a program. I want somebody else to do it, but I also want them to do it in a particular way, so they don't actually contaminate. Don't go to this district and so on, and that is a constraint in general, but I think that the benefits of actually sort of complying actually high enough, and I think for everybody involved.

Erik Wibbels:	I would just say that tablets are really cheap these days, and software to program them is free in many cases and that can be for ... One of the things we really care about is fidelity of implementation, which is about how the implementer is monitoring itself. In many cases, that can be done so much better without paper and pencil. And if it's digital, then we get to look at it too and the whole thing tightens up at I think no additional cost.
[00:36:30]	
Morgan Holmes:	Can I add one thing in. If I understood the question more to mean, "Is this whole process beneficial also to the implementing partners?"
Rebecca Wolfe:	Well, and to the Mission during the process, so one of the pieces of feedback I get from our teams working on these randomized control trials, so we're doing one in the middle bell of Nigeria, and the team is like frustrated by it, but they tell themselves every time they're frustrated, "Okay, we know we're contributing to larger learning." And they're almost at the finish line and so that's what's keeping them going the next four months. But the program will be over, the learning won't benefit that team and so they're in a sense being constrained for a public good. And so fortunately this team, I've worked with enough that they're willing to do it. Other teams, I have a harder time with, and so that's I think something we have to figure out, this incentive issue. If it's just about something, a larger policy issue, if it's about larger learning, then the team feels like, "Why am I doing it?"
[00:37:00]	
[00:37:30]	
Morgan Holmes:	Two things on that, immediately on that particular team, luckily unlike the MCC, which might be a model that people are more familiar with where we do our baseline, a huge program runs for five years, and then we do an endline. I meant to say earlier, our IEs are really more about learning, testing questions and learning than accountability. And luckily our smart PIs are creating a test that can be completed in 18 months, 24 months if it's right around an election. Sometimes it's six months, before and after an election. So we are, actually in almost all cases, able to take the findings from the impact evaluation at about year two of the program and feed that back in for the next possibly two or three years. But I also wanted to say that for the USAID's side of things, hopefully the end results do help our implementing partners to go on to replicate these programs elsewhere, but for the USAID's side, I think because of that clinic model, we're actually helping the missions with program design. Not only bringing evidence to bear from academia, but forcing them to challenge themselves in their theories of change.
[00:38:00]	
[00:38:30]	
[00:39:00]	
	Like, why do you really expect A to lead to B to lead to C? Is that really logical? Have we seen that elsewhere? Maybe we should just focus on the A to B. So even in the clinic, I think and I'm sad that Niels just walked out 'cause could have put him on the spot to answer, but even in the clinic, I think it's already adding value in how they design the program.
[00:39:30]	
Rebecca Wolfe:	Well, and so one thing I also wanted to underscore what you said earlier in your introduction about it doesn't always the whole program, and I think a lot of people think when you're doing it in RCT for an evaluation that it has to be the whole intervention, but you can have some treatment arms, and I think that's actually an

interesting part of the DRG model. Do you have something to add to that?

Chris Fariss:

[00:40:00]

For the rule of law program in Haiti, there were actually four different components of the program and at the teaching and learning clinic we considered all four and we thought, "Can we build an evaluation for each one of these?" And we ended up focusing on the one, but we went through a lot of different ideas during that week and ultimately settled on the program we did but, there was a lot of creative thinking about other programs that we could work with.

Rebecca Wolfe:

[00:40:30]

Chris, just to follow-up on that is in that week long clinic, how much are you bringing in academic research to help with the decision on those treatment arms, and how much are we bringing evidence into that initial stage of program design and testing, because I mean, there's an interest from the academic side of seeing like, "Are our theories actually working?"

Chris Fariss:
[00:41:00]

[00:41:30]

In my experience, the academics, we came in without really a very clear idea of the expectations or the goals of each of the programs. We had some phone calls. We knew about sort of the broad topography of the program, but we didn't know exactly what was gonna happen and that crystallized for I think each team in the context of the teaching and learning clinic. So from my perspective, and I think the other academics as well, there were seven teams. There were about 15 academics and then four or five people from seven different missions, so it's about 40 or 65 people, and we were all together for that four or five days experience so we learned about the country context for seven places. We learned about what was possible, and for the academics, we thought about what particular theories in political science, in sociology and economics can we link what we're learning in these programs back to.

[00:42:00]

So we didn't really know what we're gonna get when we were going in there, but we learned a lot, and so I think the academics benefited, maybe even more than ... I mean, I think that's one of my favorite experiences was doing that program and learning that I achieved myself.

Rebecca Wolfe:

And then if Erik or James, you have anything to add for the auditing experiments?

James H.:

[00:42:30]

So I haven't participated in the clinics, in this particular clinic, but I've participated in similar clinics. I think the World Bank has done similar things. And I think there isn't a greater opportunity to actually bring a lot of the research into essentially kind of that discussion. And I think that maybe an area where it will be useful to actually try to create a review of some kind that can be shared with folks before they come, because the one thing you don't want to be, at least in my experience, I think I've been most successful when I keep a lot of the research away from these conversations and just sort of try and distill a particular program, particular set of questions.

[00:43:00]

If you start rattling off the evidence ... Yeah. But I think a lot of these guys on the ground, I think they can get turned off. So I certainly think there maybe a useful way to bring a review of kind of the literature and the evidence to these guys

	before the clinic, and then I think it can actually make for a more useful interaction.
Erik Wibbels: [00:43:30]	I've had the same experience, which is we have the advantage of having read a ton of papers, probably too many papers and that means we have a sense of the weight of evidence. And so, if there's a discussion about one potential treatment, we have a sense of whether or not there's some big effects out there and some nulls or it's a bunch of ... Well we don't have a lot of cases where we just have lots of large effects.
Rebecca Wolfe:	Yet.
Erik Wibbels:	Yet, but I found it better to be not very academic even if my academic processor is using what I do.
Rebecca Wolfe:	So it's-
Erik Wibbels: [00:44:00]	One other. I do think that there is a real tension here, though in a sense that, like we want to do cool new things and it's hard to ... you know. That's a really different goal than wanting to do something that is very likely to work. We don't know that it's gonna work, which why we are gonna go test it, but I do think that this is a tension that's inherent in the overlap and that's sort of what makes it kind of fun. It's figuring out where that overlap is, but it is a tension.
Morgan Holmes: [00:44:30]	I just have to have one point of clarification then I know we'll get to Q&A, that we do actually at the clinics, 'cause we haven't had the pleasure of having you guys at the clinic yet, have the academics deliver an evidence review. Someday we should rethink it, I don't know.
Erik Wibbels:	I wrote an evidence review.
Morgan Holmes: [00:45:00]	You did write one? Okay, yeah, but so now we do have the academics and I think actually we all find it really helpful and then they refer to it later when they go home to their mission. This is generally what we know about how information campaigns have worked in different places in this regard in providing ... I think that's one of the most important components of the clinic, providing what we know so far academically.
Rebecca Wolfe:	Okay, thanks Morgan. So I think we have about 10 minutes left, so are there any questions. It's already quite a few. So how about we take three at a time? And so, yeah, I'll just point or ...
[00:45:30] Carrie Bruce:	Thank you. Carrie Bruce from Social Impact. I just wanted to ask, I think, a great presentation, but I wanted to ask the academics, how you think we should be partnering more with the contractors who are doing the same kind of work? I was especially taken, Erik, by your comments that there's so much evidence on how we ask questions that sometimes contractors who are like grinding it out in terms of putting together evidence don't have time to reflect on maybe how we could do

that better.

Rebecca Wolfe: I think there is one right behind.

Winston Allen: I'm Winston Allen. I'm with the USAID. I just want to go back to what Morgan was saying about the importance of the whole impact evaluation in form and design, and then reflecting on the USAID's evaluation on policy, which outlines when do we
actually do impact evaluation, which is for a pilot or when we're trying to test a new hypothesis. So then, just listening to the academics I was wondering to what extent these projects that you evaluated actually came with a new hypothesis or you actually searching for the theory behind it. So I'm just trying to see if it's a new hypothesis your trying to understand where are these actually work and then you scale up as [inaudible 00:46:55] or is it the other way around? I was curious how it's
going in terms of the project design and your own academic interest.

Rebecca Wolfe: Why don't we take one over here, and then we'll do the next round on this side. Yeah.

Daniel Berger: Hi I'm Daniel Berger from Creative Associates. On the implementers side I just move to the side after years in the academia, and one thing I want to say is, the
academics, we have a ton of data and we want to share it, we're actually recently having discussion about how do we let you know what we have other than just emailing our friends? So that's kinda a question on this end, and to the USAID end, we want to be inputting RCTs and small RCTs into our programs, but we're feeling like we are limited by the way the budgets are set out and it's gonna cost money to do that. We'll be serving fewer beneficiaries, not get the program and not do
anything, are we wrong? If we're not wrong, how do we tell you that we want to do this and actually still possibly be able to do something?

Rebecca Wolfe: Okay. I think we probably need the mic back up. Morgan, yeah. Morgan maybe you'll start and then ...

Morgan Holmes: I should have been taking notes on the questions.

Rebecca Wolfe: So I think Winston's question was probably more into you than those last piece. Yes.

Morgan Holmes: Okay, just to Winston's question. That's a great question. I think maybe just to
speak super bluntly other sectors in USAID might have more of a luxury to only test new hypotheses, like health and education who've been doing this longer, but in DRG we've just tested very few of our hypotheses empirically, like rigorously. And so, my feeling is and might not be exactly in accordance with the policies but we should ...

Erik Wibbels: [crosstalk 00:48:50]

Morgan Holmes: That we should test really all the development hypotheses that we're

[00:49:00]	implementing and we are now thanks to the academics again who have been doing this much more in their own field experiments under their own steam coming up with now some findings and things that maybe we don't need to test anymore in the future, but I just think that DRG is still pretty early in the curve and we still have a lot to test. That's my own personal feeling. I'm so sorry, I just can't answer that question, like I wish every USAID RFP put money in so that everybody could do a RCT. [inaudible 00:49:32].
[00:49:30] Erik Wibbels:	I'm now in this weird part of life where I read lots of RFPs and the RFPs are getting better, but they're getting better in a weird way. Where like there is someone on the team that is putting together the RFP who has heard about the stuff that we're talking about and wants to do that, but it's really unclear where that person is in the hierarchy of decision making. And so because I work with implementers on
[00:50:00]	their proposals, with the idea being that we want to build something impressive and rigorous into the proposal with the aspiration that that's gonna help it get funded, the fact that there's ambiguity there, there's some talk about learning, there's some talk about evaluation, the fact that it's also in a RFP that seems to want a 126 interventions in four counties makes it really heard to know. The first
[00:50:30]	difference is really good, but it's a confusing moment I think for the sort of thing that you're talking about.
Rebecca Wolfe:	And I can add a little bit to that because as Mercy Corps, we've done some RCTs that it wasn't the decision of the Mission, we decided to do it. So the one I was talking about in the middle bell, for people in the audience, it was a conflict
[00:51:00]	management mitigation program. This is funded every year and we decided this has done so much, we needed to do a more rigorous evaluation, and I actually then went down to the Mission and we got more funding for it. So it's also how do you then renegotiate this once the contract is awarded. Because often these are appeased or written also by committee, so you have multiple interests, but then there is a team on the ground actually managing the contract. That's a good place to do the negotiation.
[00:51:30]	I think there were Carren's question and then I guess the other part, so Chris, yeah.
Chris Fariss:	On the contractor's question. I noticed in minor action with the large team, with all the members from different implementers and the academics that sometimes there'd be different disagreements, but the different disagreements never involved more than two of those actors. And occasionally there'd be a disagreement about like the survey design. And the survey design disagreement would only take place
[00:52:00]	between the people who are gonna implement the survey and who are actually tasked with creating it and implementing it and then, me and my co-author as the academics, that we would disagree about little details but then that disagreement happened and we sort of figured out what the best choice was and then we more forward.
	And then there was disagreements about other stuff that I'm not ever gonna try to tell you about 'cause I don't remember, because at that point I was like, "Well this

[00:52:30]	is not my point to be talking in the phone call." But I think having two, at least two people that could argue out the issue and come up with the best decision a that moment was useful, just from my experience with the contractor, so I think the Haiti project was helped by the fact that there were a couple of different groups that had an incentive to talk about the survey and the implementation of the design.
[00:53:00]	On the data issue, I think there's a lot of academic infrastructure for data housing and data sharing and some of those they just come off the top of my mind, top of my head, is the Dataverse at Harvard and ICPSR in the University of Michigan where I'm on faculty. And there are resources for engaging with those groups, so that might be one way to do that.
[00:53:30] [00:54:00]	And then the final thing I wanted to say, and I don't know which with this was about. It was about RCT implementation, maybe without budget or in future programs, once there's evidence about a particular invention. For the Haiti project, if this project were to be implemented again without funding for an impact evaluation component. There's no reason that the randomization of ordering of services couldn't be made part of that at the beginning. And there's no cost to that at all then by randomizing the order that the intervention is distributed over the clients, we get the benefit of being able to evaluate at every moment that program's running, whether it's effective, or if the ethicacy rate is changing, or if it's different in different jurisdictions or in different countries. And so that randomization component can just be part of the program going forward and I hope is.
Rebecca Wolfe:	Okay, James or Erik, do you have anything to add? Or why don't we take a couple more questions. We probably have time for two more.
Mark Koenig: [00:54:30] [00:55:00] [00:55:30]	Mark Koenig, transitioning from USAID Congo to USAID Guinea. This has been a really interesting conversation and a lot of it kinda dancing around what are the local incentives and who are the local actors and how do they play in the decision making, and I thought I would toss out a couple of different possible interested parties and sort of get a sense from you of who are the most active? And then also importantly, the capacity building locally, because it's not simply that you're doing this research activity, but people should have an incentive to learn how to do this because it will help them in future projects. And so, within a mission, it's likely to be the technical office involved in whatever your measuring, DRG or other. Program office is likely to be involved. Within the mission also you'll have your short term ex-pats, but also you have long term local hire, what we call FSNs, some of whom are very highly qualified and so I'm kinda interested on the mission side, who do you work with in the mission and where do you find the best long term promise for a leave behind in terms of future capacity?
	And of course you have your implementing partners. You also have your very interesting Imine mechanisms, emissions will have mechanisms, and they of course will have an interest in this. So I just tossed that out as sort of an additional area to explore.

Rebecca Wolfe: Okay, thank you. One more question.

Neetha: Neetha, from IRI. Actually that's a perfect question and I wanted to build on it. Are
[00:56:00]   we doing enough to build the capacity of in-country universities and researchers,
                      not to take away from US business, and understanding the role the longitudinal
                      impact plays in all of this. Once the program is closed, are we revisiting that
                      information and how are we revisiting it and how does it inform future program
                      design. And then quickly I want to throw in one last, as implementers in the room,
                      we all do our own evaluations, we all do rigorous evaluations and I knows there's
[00:56:30]   the DAC, but are we doing enough in the space to share and make sure that we're
                      just not reinvesting the wheel or doing the same programs over and over again?

Rebecca Wolfe: Okay, who wants to start? Okay.

Morgan Holmes: [inaudible 00:56:46]

[00:57:00]
James H.:   Again, hot potato. So, I think that's a really good question. In a lot of the places that
                      I work, I work mostly in Eastern Southern Africa, it's always a mix bag whether in
                      fact you're gonna find the kind of researchers who are gonna be interested in doing
                      this work. In some places I've had some luck in finding collaborators. Rwanda is a
                      great example where in fact people, in fact the government almost structures it so
                      that in fact you have to work with local actors. In Uganda right now actually, we
[00:57:30]   have to have a local PI as part of kinda the review process, you actually sort of find
                      somebody who will do this. But I think the speed of which these things move, just
                      makes it really hard sometimes to always go out find the people to do this. But I
                      should say one thing, a lot of the data collection, a lot of the analysis, a lot of the
                      RAs, those tend to be hired locally, and I think there's an incredible amount of
                      capacity being generated, and IPA is probably kinda an interesting example where
                      they've hired all these guys, trained them in a lot of these measurement and
[00:58:00]   analysis and they're now actually setting up their own firms and actually even
                      challenging IPA for contract.

                      So I think there is something, but it's probably not going as far as it should.

Rebecca Wolfe: I recognize I did not manage time particularly well, so if any people have short final
                      comments? Yeah.

Erik Wibbels: On the university side we've had great academic partners in Ghana. On the data
[00:58:30]   collection side, all over the place, so it's been a great success. This is a really
                      complicated thing, I just want people that took the blue pill, and it's hard to know
                      ex-ante who took ... I mean there's the hard grinding work of getting to agreement
                      on what's interesting and worth doing and rigorous, and then you still need the
                      blue pill takers and I don't ... I don't know, having tried this a lot of times now, I
                      don't know. I mean I know agreeable types and disagreeable types, but that's all.
[00:59:00]
Morgan Holmes: Our foreign service nationals like you have Emanuel in Ghana who's been an

amazing ... yeah. The foreign service officers rotate out every two or so years, so for most of our IEs, we really rely on the foreign service nationals who have been our most steadfast partners.

Chris Fariss: [inaudible 00:59:22]

Rebecca Wolfe: Okay, well thanks Chris.

Chris Fariss: Thanks for having me.

[00:59:30]
Rebecca Wolfe: So thanks everyone for coming. Obviously there's a lot more to talk about and hopefully we'll find other ways to engage this conversation. Thank you.