

Daniel Handel: Good evening everybody. Welcome to the first session of a very good day. My name is Daniel Handel. I'm a Senior Advisor on Aid Effectiveness here at USAID. I am particularly interested in this session as I am a foreign service officer, so I am one of those guys tasked with trying to spend our money well. I have personally benefited a lot from 3ie and their work and their website, which we were just discussing.

[00:00:30]

Let me introduce Hugh Waddington and Anna Heard who are senior evaluation specialists at 3ie. Take it away.

Hugh  
Waddington:

Thank you very much. We appear to be missing Anna but I'm sure she will be here in 30 seconds or so. Thank you very much for coming to this session and to USAID for supporting us. We're going to try to be a bit interactive, so there will be slides and stuff. But then we actually have a good number here to do two breakout sessions, so there will be two breakout sessions coming up quite soon, so you won't just be listening to me and Anna talk the whole time.

[00:01:00]

This is all about how to commission and use systematic evidence, so that's systematic reviews, we will touch on evidence gap maps and we will talk about statistical meta-analysis as well. For those of you that don't know 3ie, you can think of us as an evidence intermediary. We help organizations conduct rigorous studies, whether that is a rigorous primary study like a randomized control trial or a quasi experimental study, or to do rigorous synthesis work like systematic reviews or evidence summaries like evidence gap maps.

[00:01:30]

[00:02:00]

These are the partners that we have worked with. We're proud to have worked with these people and we gratefully acknowledge support. We do have a ... You'll have noticed on the chairs in front of you these plain language summaries, many of which are from the Campbell Collaboration. 3ie, we partner with the Campbell Collaboration in the production of systematic review technical work. The Campbell Collaboration is the ... Those of you that know that Cochrane Collaboration, it's the [inaudible 00:02:26] Organization of Cochrane, which is devoted to social and economic interventions.

[00:02:30]

So, this is our three broad objectives. The first bit, it's broadly a kind of indoctrination for the first maybe 10 minutes or so, which is about what, why and where to find systematic evidence. Then we're going to work on these two latter parts on how to ask answerable questions and what processes are needed to produce systematic evidence. Both two crucial components in getting you, I'm actually going to ask you in a second, but you I'm assuming is a room full of policy and practice people, how to commission a review that gets you the evidence you need, answer the questions you want, in the timeframe you want.

[00:03:00]

[00:03:30]

Perhaps first here we can have a quick show of hands people that work in policy. Practice like more implementation side? Research? So, some people doing a little bit of moonlighting, okay. We wanted to ask you, if someone asked you to prepare a brief in three hours on the effectiveness of a particular intervention, where would

[00:04:00] you go for sound evidence? Where do you go? This is an open question for those of you out here. Let's start with the research people. Let's put the research people on the spot.

Speaker 3: Publications.

Hugh Waddington: Any particular organization or repository?

Speaker 3: I like-

Hugh Waddington: Oh, we need to use the mic.

Speaker 3: I use Google Scholar a lot to search publications.

Hugh Waddington: Google Scholar. There you go.

[00:04:30]

Speaker 3: Yeah, and then of course from peer review journals usually we'll take those first as we're looking for different publications and literature.

Male: That's a lot in three hours.

Speaker 3: In my particular field of work, there is a lot of ... I'm sorry, you don't need to hold that for me.

Hugh Waddington: The mic holder. He's an employee.

Speaker 3: In my particular field of work, there's a lot of websites and different resources that actually pull on systematic reviews, they'll pull on meta-analyses and so I would go to those for my particular field.  
[00:05:00]

Hugh Waddington: Anyone from the more policy or implementation side?

Speaker 5: There's a lot of gray literature out there, and while it may not be as rigorous as some of the peer review journals, there is increasingly some good evidence out there I think from the practitioner world as well.

[00:05:30]

Hugh Waddington: Right, and any particular kind of databases ... Any particular repositories you use, like the USAID has a website does it, with a database or ... Okay, so basically, it's a hard task to do this if someone asked you to do it. And, you'd probably have your preferred source of evidence. I'm sure there's USAID databases.

[00:06:00] But the problem is, then there's all of this, and you've talked about this. There's all these organizations doing rigorous impact evaluations like 3ie, World Bank DIME, JPAL, IPA and so on. And then there's how many hundreds of thousands of journals, gray literature was mentioned. There's just so much out there that it is literally beyond the capacity of the human mind. So, we need to delegate this task to what

- [00:06:30] we like to say, intelligence researchers or intelligent machines to do this work for us.
- Just a little bit of backtracking, why would we use a single study? I'm thinking of single studies like RCTs. A single study in a particular context can be extremely useful. It can help us scale up what works. A study we funded in Mozambique on early child development, which found very big impacts both on the children in terms of their nutritional and cognitive development, but also their siblings as well, who were released to be able to go to school.
- [00:07:00]
- This program was scaled up in Mozambique. From this study we've done rigorous randomized control trial. Or perhaps we might close down or redesign one. Here's a nother study which was done in Northern Ghana on cook stoves. The program didn't work because it was largely polygamous communities and it was very difficult to situate the cook stove within one particular household. The other wives didn't particularly like that. We can use them to inform policy discussions in that particular context. We've funded studies at the youth wage subsidy in South Africa.
- [00:07:30]
- The problem is that single studies are biased when we're trying to get a sense of the overall evidence. Why are they biased? They don't tell us about what the balance of the evidence is in different contexts or at different time periods. They tell us only about one part of a policy issue. That might be just the impact question or it might be the cost effectiveness question. It won't be a thorough as we might desire.
- [00:08:00]
- They are specific to the particular samples of people that were drawn on in conducting the study, the time period and so on. And they're often of poor quality. Many single studies that are published in peer review journals are of poor quality, and many gray literature studies are simply not published in journals, because they don't find an effect. They find a null effect, and that isn't considered to be interesting enough to get it published in a journal. This is a major problem with drawing on published sources of rigorous evidence.
- [00:08:30]
- [00:09:00] Here's a study that we did on farmer field schools. Farmer field schools is an adult education approach really, although some people think of it as agriculture extension. But essentially, in a farmer field school, a group of 25 farmers are taken through different type of agricultural practices over the entire course of a growing season by a facilitator, so it's very time intensive.
- [00:09:30] You can think of it as time intensive extension, or you can think of it as adult education. The objective of these approaches are either to get farmers to use more organic type of practices like for example, integrated pest management, or to improve their agricultural outcomes or to empower them to be able to think creatively to solve problems.
- [00:10:00] There was a study done by the World Bank. This is a single impact evaluation. It wasn't an RCT, but it was a rigorous impact evaluation that concluded that of this intervention undertaken in Indonesia, there were no effects. So, there were no

effects on adoption of practices, there were no effects on yields, and there was no effects on spillover effects to non-participating farmers in the community.

[00:10:30] On the other hand, the FAO who actually developed this approach back in the late 1980s, they came out with a review a couple of years later that said that wide spread, remarkable and lasting developmental impact. So, who do you believe? You have these two sides. And the single study by the World Bank incidentally was cited in the World Bank document where they pulled out of the global IPM facility, so it did have a fairly substantial policy impact.

[00:11:00] We would say that actually in order to know what the evidence says, you need to do a systematic review. You need to do the systematic collection, critical appraisal, and synthesis of evidence if you're going to say something generalizable for policy about lots of contexts. This is the summary results of a meta-analysis. A statistical meta-analysis enables you to pull impact findings from lots of different contexts.

[00:11:30] What did we find? That in small scale pilot settings, on average, typically programs found a 25% reduction in pesticides. That was one of the positive objectives of the schemes. A 10% increase in yields, or a 17% increase in income. That's a substantial ... You can think about your income being increased by 10%. That's a substantial increase.

[00:12:00] However, this was ... Sorry for the slide that you can't really read properly. These were pilot programs. The problem with farmer field schools is that the facilitation process needs to be done very well, needs to be targeted, delivered, and followed up and backstopped appropriately. And so, when you try to scale up these approaches at a national level as they did in Indonesia, we didn't see the effects anymore.

[00:12:30] Our conclusion was that it's just simply difficult of impossible to recruit, train, and keep farmer field school facilitators at the national level going around and being able to facilitate these programs.

[00:13:00] So, what ar systematic reviews? Essentially, it's a way of balancing what the evidence says. And the evidence may be on impacts, it may be on targeting, it may be on cost effectiveness and so on. Its kind of an accountability exercise about the evidence base as well, so it helps us know what evidence is actually rigorous and what evidence is not. And it helps us to understand what we can generalize and what evidence is rather more context specific. As I've mentioned, it's also very useful in addressing some really important problems in published evidence such as publication bias.

[00:13:30] Here's a quote from our 3ie's new global director for stuff. A very important person in 3ie. "Drawing on systematic critically appraised evidence is a no brainer," says Marie Gaarder form 3ie. What you can see on this slide if you haven't seen one before is what we call a forest plot. Has anyone seen a forest plot before? Just put your hand up if you have. Two people, okay.

[00:14:00]

[00:14:30] Essentially, what a forest plot is, it enables us to connect together lots of individual studies conducted in lots of different contexts, and present a summary figure about what the impacts are. In this forest plot, the studies are looking at the impacts of formal land titling, so that's the intervention, these are the individual studies. The outcome that they're looking at is land's productivity.

[00:15:00] All of these individual studies have a ... Let me use the pointer. Have a point estimate. This is the impact in the intervention group over the impact in the comparison or the control group. And that's always associated with a 95% compliments interval. This is the likely impact of formal land titling on land productivity in Nicaragua.

[00:15:30] What you can see it that there's all of these studies, and these officers, Steve Lawry of the World Agroforestry [inaudible 00:15:42] in Indonesia and Cyrus Samii, they conducted this synthesis. And what did they conclude from the synthesis? Well, land titling is certainly effective in some instances, but when we separate out the studies by global region, what we're seeing is that actually we see much bigger effects in Asia and in Latin America than we do in Africa.

[00:16:00] There's some very good reasons why we might expect that to be the case, one of which being that communal land tenure is stronger in some countries in Africa who affords more support for ... Reduces risk of farmers there. You wouldn't expect formal land titling to have a bigger effect, but also of course, farmers in Africa lack access to the types of upstream and downstream inputs and access to markets, which might make a difference to investment.

[00:16:30] So, what makes a review systematic? Systematic searching for studies, generally like when you do a good quality randomized control trial or impact evaluation, you'll be required to publish a pre-analysis plan, so a protocol is like a pre-analysis plan for a systematic review. It says what the searching is going to be and what the inclusion criteria are. So, critical appraisal to what we call, separate the wheat from the chaff, separate the higher quality from the lower quality evidence.

[00:17:00] Systematic and transparent data extraction, analysis, possibly meta-analysis, possibly synthesis of qualitative evidence and systematic reporting of findings and implications. So, the kind of ... If you were to ever read a systematic review report ... I'm not sure I should be saying this. It's not advisable. There are generally 400 pages of, "Our process was this, and this is everything we did. These are all our findings."

[00:17:30] It really is not a user friendly document at all, but we produced our user friendly summaries and these plain language summaries Campbell helped us produce to get the messages out more clearly. Outside you'll have seen 3ie produces our systematic review summary reports, which kind of ... Longer than a ... They used to be 25 pages. Some might be a bit longer, but they try to give more contextualized evidence than you can do in a one page or a two page brief. And systematic engagement. This is part of our engagement, but we'll talk more about that later.

[00:18:00]

[00:18:30] Sources of systematic evidence. These are some of the organizations that produce systematic evidence. We have our partner the Campbell Collaboration. The original collaboration being Cochrane that publishes reviews in the area of biomedicine and public health, and also healthcare. Then there's Collaboration for Environmental Evidence, which is based in the UK, which focuses on climate change in environments. And then The Epicenter, which is a supporter of authors doing reviews and also publishes a library.

[00:19:00] So, there's all these libraries already for systematic reviews. I would say, if you were to try to find systematic reviews, you could do a lot worse than going to 3ie's database. 3ie's database, we've just recently updated it so there's now over 700 records, and includes both a summary and a critical appraisal of the reviews as well. And that's 700 records.

[00:19:30] When we started doing this back in 2008, I hadn't even heard of systematic reviews. We've just seen a massive explosion. It's like the kind of explosion or RCTs that's been a very, very big increase. There's potentially a lot of reviews out there that you can draw on. They might be out of date, but you could potentially draw on them.

[00:20:00] Now, we're gonna get to the first exercise. I'm just gonna take you through some concepts, and then we're gonna get to the first exercise. This is really about how to ask answerable questions. The systematic review process is very ... It's like doing a primary study. When you go and do a primary study, you go to the field, you will do piloting, you will do data collection and you will do checking like double coding to make sure the data was collected accurately.

[00:20:30] Then you will do your fancy econometric or whatever type of analysis with a baseline and an end line. Then you will produce a report and maybe you'll see an impact some years down the line. That is like what a systematic review is. A systematic review is that different from doing a literature review. Just to be clear, a systematic review, we are drawing on literature which is already out there, so we don't generally go to the field to collect additional data.

[00:21:00] We might contact organizations that either run programs or authors that have produce these evidence to obtain additional evidence about the studies, but essentially this is a desk based piece of work. And it can be used ... Oh yes. Yes please but, can we give you a microphone? No you can't. I've been told you're not allowed.

Victoria: Thank you. I'm Victoria [inaudible 00:21:49] Company. We do economic growth in fragile states. One question I had, you just kind of mentioned it is the quality of data. We do for example a lot of work in Iraq and Afghanistan where we are not the primary of the data. We relying on local partners. There's been lots of folks on the ground to understand that sometimes that data is not collected appropriately.

[00:22:00] Is there a way that you account for this type of problem when you approach a systematic review in terms of understanding or validating the quality of the data

that you're working with in the first place?

Hugh  
Waddington:

Yes, absolutely. Systematic critical appraisal, it's ... Any information that is included in a review is critically appraised. And again, it's like this kind of internal quality assurance. There will be two authors that do that and possibly a third author to kind on invigilate basically. The quality assurance is built in both internally and externally into the process.

[00:22:30]

[00:23:00]

So, there's different types of evaluation questions. I'm sorry for this. I hope it's the only technical language I'm going to use. We can think of like first generation types of impact evaluation questions which is, does an intervention work as opposed to not doing anything else? Just, does it work? Does it actually make a difference? Does it really work?

[00:23:30]

Then there are second generation types of questions, which are often more relevant for people actually taking decisions about what to implement for example. These questions are more about, does intervention A work compared to B, C, D and so on? So, really a kind of relative effectiveness type of question. Similar to that would be what different intensities of implementation are most effective and, maybe even different implementers, different approaches to implementation.

[00:24:00]

These are broadly categorized into first and second generation questions as I'm calling them.

[00:24:30]

These correspond to two main types of reviews. We have a review by Carinne Brody and Thomas de Hoop from AIR on the impacts of women's self-help groups. This is groups that bringing together women usually in rural areas, may be providing some financial supports whether it's savings or micro-savings or credit. And usually, also providing some kind of additional messaging, maybe health messaging or something like this.

[00:25:00]

This is a single review of a single intervention, which looked at lots of different evaluation sub-questions within that intervention. It looked at not just what is the impact of the intervention as opposed to not being in a self-help group, but also, how were self-help groups targeted? Was targeting effective? What were participant views about what they got out of the self-help group?

[00:25:30]

Then there's a second type of review question, which is the more relative effectiveness type question. What works best in particular contexts in improving school attendance and test scores? We have a very nice, recently published summary report of this. I'm not sure if there are any copies outside, but they're on the stand. Sector-wide review covering most or anything you can think of in improving school outcomes.

[00:26:00]

So, two broad types of reviews. And how do we turn a question that we want answered as development practitioners, into something which is answerable in the systematic review, given the rigors of doing a systematic review and what it entails? We use this acronym called PICO. PICO acronym, you can also apply it to impact evaluations, you can apply it to many different types of research.

- [00:26:30] Essentially, it's population in terms of what is eligible for inclusion in this study. Populations from which region, or perhaps countries, or perhaps locations within countries; rural or urban, or perhaps which particular types of groups of people. What does the evidence have to cover in order to make it into a review?
- Intervention, is it the single intervention like self-help groups or is it just anything like what works in education? The comparison, so, what are we actually comparing? What is our ... When we're calculating the impact, what are we comparing the effectiveness of our intervention to? Is it compared to doing nothing or perhaps a more standard approach?
- [00:27:00] The outcome. Generally, a review of ... Most reviews will specify particular outcomes that are of interest. So, a review will need to search for evidence on particular interventions and it will either incorporate any outcome or relevance, or
- [00:27:30] it will specify particular outcomes and perhaps search on those outcomes. So, 3ie's review on education effectiveness looked at the primary outcomes for school enrollments and attendance and test scores, but it also included additional secondary outcomes.
- [00:28:00] There's something that I don't really want us to go in, because I don't think we have time. The S in PICO if we turn it into PICOS is about study design. That is about what type of evidence is appropriate to answer a particular question. Sometimes the best evidence is from an RCT or a quasi experiment and impact valuation. Sometimes that evidence is from a qualitative study or an ethnography or a participatory type of study. It all depends on the question.
- [00:28:30] Just quickly, we're almost at the exercise. What we tend to do is ... It's okay. Small amount of attrition is fine, I can handle it. What we tend to do is, we work with decision makers to commission reviews that answer the types of questions that you wanna answer. And we have a consultative process whereby we go about doing that.
- [00:29:00] We consult, we conduct a needs assessment, which might be something like an evidence gap map, or at least will be an assessment of what the available evidence is for primary studies and for systematic reviews. We work with you to clarify the question and then agree it, and then we PICO it up and put it into a call for proposals essentially.
- [00:29:30] An example of a PICO'd question; do public schools that receive computer assisted learning get better test scores than public schools which in low middle income countries? That's an example of a question, which is eminently answerable in a systematic review, and indeed is one of the sub-questions of this broader sector-wide systematic review.
- This is an example of an evidence gap map. If you to 3ie's stand, you'll be able to browse. Marie got some nice functionalist with her computer and stuff. A gap map, it's like the first step in doing systematic review. We developed it to provide results

[00:30:00] from searches more quickly than a systematic review can do.

[00:30:30] It's essentially a matrix of interventions and outcomes along the causal chain, and then basically a bubble chart showing the distribution, density and porosity of evidence so, where's the evidence, where's the gaps? Here's a gap map we used in wash in order to commission some primary impact evaluations and some systematic reviews that looked at the outcomes that were lesser reported.

[00:31:00] We're almost at the exercise. We can probably think of four main types of systematic reviews that we do and I've already mentioned two of them. We have a review of effects, which would be a single intervention, multiple outcomes. An example being, what is the impact of daycare on outcomes for children? A range of outcomes. Comparative review of effects, so, the impacts of education programs in general.

[00:31:30] There are reviews of economic evidence so, cost effectiveness or willingness to pay and so on. And then what we call an effectiveness review, or a full causal chain synthesis, which is a review like the one we did on farmer field school where we focus on a particular intervention, but we answer lots and lots of detailed questions using different types of evidence.

[00:32:00] Those are the broad ... You can either go broad but have a more specific focus on a particular question, or you can go narrow on the intervention, but then be able to answer potentially lots of questions that are policy relevant. Okay, over to you.

[00:32:00] Anna Heard: So, what we wanted to do or the idea was to talk about some types of different questions that you might want to answer in general in development. If you could answer a question by doing some research, what might that question look like? Who has a question about development that they'd like answered? Any question. Yes.

Daniel Handel: Of the many drivers-

[00:32:30] Anna Heard: No, I'm just asking you to write it in that.

Hugh Waddington: Oh right, okay. You'll have to hand over the mic then.

Daniel Handel: My previous country was Rwanda where malnutrition is a big issue. Of the many things which could cause malnutrition, are we better off programming money that does a little bit of everything, or should we try to focus on one or two constraints? Or main drivers [inaudible 00:32:55]?

[00:33:00] Hugh Waddington: That's like, what are the main drivers of ...

Anna Heard: How do we focus funding for malnutrition?

Daniel Handel: That's a better way to say it.

Anna Heard: Anybody else have a question? You have a question? No. You have a question?

Speaker 5:  
[00:33:30] There's been a lot of attention recently to unconditional cash transfers as being potentially a more cost effective approach compared to traditional development, more wholistic approaches that involve behavior change. But there's been very little focus on the sustainability of unconditional cash transfer, sustainability of the impact of those versus a wholistic approach.

Anna Heard: So, what are the longterm effects of unconditional cash transfers as opposed to more traditional approaches?

Hugh  
Waddington:  
[00:34:00] We will bring the screen back.

Anna Heard: What are the longterm effects of unconditional cash transfers versus other approaches? Who else has a question?

Speaker 8:  
[00:34:30] I have a question about evidence informed policy making. By that I mean trying to increase the use of data and evidence when policy makers set priorities in development. The question is, are there interventions that are proven to be more effective, giving trainings for policy makers on systematic reviews versus trying to write legislation, mandating these of systematic reviews? Putting an office in place, a chief evaluation officer where this is his or her role so what kinds of interventions to increase data and evidence use in policy making?

Victoria:  
[00:35:00] Again, kind of back to our experience in Iraq and Afghanistan. We do some entrepreneurship work. The question I have is, do women in these countries actually want to start companies or work? Because what we are finding is that they actually don't a lot of times, or at least they would just prefer to have some income for two or three hours a day versus starting their own company, which seems to be the direction the funding is going.

blue suit, hair: One that we are working on right now is specifically, how would you apply metrics for defining what a systems leader looks like to create systemic change?

Hugh  
Waddington:  
[00:35:30] How to define metrics ...

blue suit, hair: For what a systems leader looks like.

Hugh  
Waddington: Could you explain that? You can explain it in the group exercise, great [crosstalk 00:35:39].

Anna Heard: Did you have one? Nope? Okay.

Hugh  
Waddington:  
Anna Heard: So that's five-  
Can we get those up on the board?

Hugh  
Waddington:  
Anna Heard:  
[00:36:00] Yeah, okay.  
Okay so, what we have now are five questions that may be more or less amenable to a systematic review. All of these can be answered by doing some additional research. Some of these are very broad, some of these are quite narrow. What we wanna do is split up into five different groups, take one of the questions. Take that question and craft it into a question that could be answerable with a systematic review. Does everybody understand the point?

[00:36:30] What did we say would be a good question for a systematic review? We used one particular system to help us take that question and make sure it is answerable with a systematic review. The framework that we're using is the PICO framework. Within your question, can you include these types of specifications, in order to make the research question more specific and answerable with some kind of systematic collection of data and a review.

[00:37:00] It's hard to have both the PICCO criteria and-

Hugh  
Waddington: You're gonna have to remember these or we will remind you when we walk around, PICO.

Anna Heard:  
[00:37:30] Population, intervention, comparison, outcome. The five questions which we now have. So, let's have one group in the back of the room over here, one group over there, one group here, one group here and then one group up front maybe and just ... I don't know, do you wanna number off or do you wanna ask people to choose a question?

Hugh  
Waddington: If you volunteered the question, I guess you'd be part of that group. So number one here.

Anna Heard: Two, three-

Hugh  
Waddington:  
Anna Heard: Three at the back on the left.  
Four, and then five up here at the front.

Hugh  
Waddington:  
Anna Heard:  
[00:38:00] Yeah, five at the front here.  
What we're gonna ask is for you to ... Within your groups, re-craft the question to make it answerable with a systematic review, and then when you're done, you're gonna report out what your ultimate question became. We're gonna start with just going quickly through the question that we came up with, and then if we have time for one or two questions, or something that you took away from this exercise that you would really like to talk about, we'll do that. Who wants to go first?

[00:38:30]

blue suit, hair: My group was massive. We actually did do a good refining of the question to come to, what kind of metrics would be best for predicting systems leadership success? So, here you go.

Anna Heard: Yeah, you ready?

Speaker 10:  
[00:39:00] We were question number four. We had a lot of discussion and it finally came down to, what are the factors that influence women's decision making process when deciding to participate in entrepreneurship programs-

Hugh  
Waddington: They're very answerable.

Speaker 10: ... in Asia?

Speaker 11:  
[00:39:30] I'm gonna go to you for the intervention part. We just did the PICO definition. We didn't formulate a question. So population, what we're talking about global policy makers. This is question three I'm sorry, question three. The population, global policy makers, policy makers around the world. The intervention ... Can you go through those three?

Speaker 8:  
[00:40:00] Basically like, if we just narrow it to systematic reviews, is it teaching policy makers what systematic reviews are to get the, to use more of them? Is it creating legislation, mandating that they use systematic reviews before they make a policy decision? Or is it actually setting money aside so that policy makers can commission systematic reviews? Those types of interventions, that's what we're talking about.

Speaker 11: The outcome is just the use of the systematic reviews in policy making, which we haven't figured out how to measure. We'd probably have to narrow the scope to a sector or something like that.

Anna Heard: That question was actually fairly well crafted for a systematic review to start with, so yeah.

[00:40:30]

Speaker 12: I don't think we made major changing to the wording of the question, but we did in looking at the PICO, focus in on Rwanda in particular. There's a study that's currently going on there that we could eventually expand upon to look more at the long term effects, and then focusing in particularly on the question of nutrition.

[00:41:00] There are lots of different approaches to dealing with malnutrition and looking at ... comparing this unconditional cash transfer, versus another package of services or perhaps a couple different types of package of services. We kind of stopped on that point. We didn't really cover that completely. Anything that I missed?

Daniel Handel: It ended up actually being a question of you guys. Interesting point is, is it best to do a study on every possible intervention, every possible combo of interventions

that seems probably prohibitive? How could we move towards better cost effectiveness in a world where we probably can't literally test everything we could ever do?

[00:41:30]

Anna Heard:

Part of that has to do with your question, or what exactly is the goal of answering your question? If you're trying to figure out of all of the different interventions out there, which one is the most effective in improving malnutrition, then you're gonna probably focus on that outcome of malnutrition. And then start looking at all of the programs out there that are looking at malnutrition and doing a systematic review of programs who have an outcome of some sort of nutritional status. It would end up being a little bit like our education review that's giant.

[00:42:00]

Hugh

Waddington:

Speaker 3:

Right, the last group please.

Because you guys decided to shift your cash question to the nutrition sector, we do have a lot of ... We came up with similar things. We didn't form a question, but we did answer the PICO. We wanted to focus in Sub-Saharan Africa, in an emergency setting as opposed to a development setting, and focus on children under five. I'm sorry, we're question number one. The malnutrition question.

[00:42:30]

[00:43:00]

In terms of intervention, we were looking at food based intervention, cash based intervention versus behavior intervention. And then in terms of outcomes, we recognize that there's a lot of different ways to measure malnutrition, so we wanted to focus on acute malnutrition, MAM and SAM. Then there's also the introduction of maybe looking at relapse as well to understand the sustainability of the intervention.

Anna Heard:

[00:43:30]

Great, these are all really good. I think the last take home message that I would like to send home with you is, keep in mind that when you're doing a systematic review, you're not trying to answer a specific question for an individual study. That you are trying to get what the body of evidence is telling you about these different areas.

Hugh

Waddington:

Anna Heard:

Thank you.

Great.

Hugh

Waddington:

[00:44:00]

Okay, so that was a surprisingly successful exercise. Doesn't always work out so well, so that's great. We have 25 minutes now. We have to finish on time. We're gonna take you through 10 more minutes of stuff, and then we're gonna have a final exercise based on these plain language summaries which are in front of you.

[00:44:30]

Let's just talk a bit about doing systematic review. Yes, please feel free to stay where you are or move about. I don't really mind. We focus on studies that are both rigorous and relevant. We do set ourselves a bit of a challenge. Often you will get one or you will get the other, but you won't get both.

3ie's principle is both rigor and relevance. So, to use the best technical methods to answer the questions that are the right questions that need answering. What does that mean? Stakeholder engagement, to ask the relevant questions, and potentially use mixed methods or different types of evidence.

[00:45:00] A synthesis approach, so possibly for impact evidence, the correct synthesis approach is meta-analysis and that because it takes into account certain types of information that other methods of synthesis don't. Systematic reporting of both findings, but also what the implications of the findings are for the readership. And then use a friendly knowledge translation.

[00:45:30] No, there's one more. Timeliness of course. How would I forget timeliness? You could imagine that doing this rigorous evidence synthesis approach can be very time consuming. It depends on how broad the team is and it requires a lot of coordination. But essentially, these are some typical timelines and costs of doing full systematic reviews.

[00:46:00] We talked about the single intervention review that just answers an impact question. You're probably looking at \$100,000 over a period of 12 to 18 months.

Speaker 13: So, this is to inform the project planning?

Hugh Waddington: This is to inform the project planning. It's okay, we've got the question, to inform the [inaudible 00:46:13]. This is essentially how long and how much it will take to do a systematic review.

Speaker 13: To inform program design?

Hugh Waddington: To inform program design.

Speaker 13: Okay.

[00:46:30] Hugh Waddington: You ought to inform information on what policies to implement and that kind of thing. So, that's a single intervention review. The more comparative review where we have lots of different types of interventions you could imagine is more burdensome. We're looking at typical costs upwards of \$100,000 over a time period of one to two years.

[00:47:00] And then we've called this full causal chain effectiveness review, which is a review that answers lots of different types of questions using evidence appropriate to answer those questions, so a mixed method review. Again, we're looking at a similar kind of amount to a two way causal review.

[00:47:30] You can see that quite often, usually actually, the policy question is, "Where is the evidence? I need it by yesterday." If we're lucky, it might be, "You've got three hours," or, "You've got a couple of weeks." So, we're really looking to build up the systematic libraries of evidence about what we know. I've told you that 3ie has

collected together a repository of over 700 systematic reviews already on development topics.

[00:48:00] When we're commissioning studies, we need to be realistic about how long these studies take to do, and whether it's a systematic review you need, or whether it's something else. But we can be intelligent about it. We can build systematic reviews into know processes like for example, the World Bank's World Development Report on Education is coming out next year that partially informs why we did the review on education, and that review on education is now feeding into that report.

[00:48:30] It's also worth adding ... Or, we commission through communities of practice. That's pulled funding with different donors and decision makers that are working towards a particular topic like wash sector for example. We commission with the Water Supply and Sanitation Collaborative Council.

[00:49:00] Reviews need to have technical support. I'm sorry to any ... I don't think there are any just pure researchers in here, but it is not a good idea to allow academics to peer review their own work. There is an external quality assurance process, and there's also a technical support which is needed for particularly new author teams because the systematic review, it's not just learning the technical aspects of doing a meta-analysis for example. It's also how to coordinate and make sure that the review uses the right processes.

[00:49:30] If you're looking at something a bit quicker, there are ways of ... First of all, a problem that we have is that a lot of reviews are out of date. We've started this practice we almost 10 years ago now. A lot of the reviews that we commissioned were maybe drawing on evidence, which is now five or six years old so these reviews need to be updated. A systematic review is only as relevant as the information it contains.

[00:50:00] So, there are updates which are starting. We are starting a program of updating reviews to make sure that ... And we're looking for partners in order to do that. You can also have things like reviews of reviews now. Like in the microfinance sector, there is something like 40 systemic reviews on everything from micro-savings to micro-credit, insurance and so on. We are commissioning a review of reviews of the evidence, broadly in the financial inclusion sector. This is the broad approximate costs and timeframes for doing that.

[00:50:30]  
Speaker 15: Quick question.

Hugh  
Waddington:

Speaker 15: For a review or reviews, is this something that we could also implement around governance projects or just [inaudible 00:50:41]?

Hugh  
Waddington: I can tell you that ... I'd mentioned evidence gaps maps. We've published a gap map on a state society relations, which is broad evidence on governance topics, which

[00:51:00] incorporates primary cities and systematic reviews. That is like the start of a review of reviews on that topic. I'm sure that we have hard copies somewhere like at our stand on something like this but yes, certainly a review or reviews, it's still not been done and probably needed.

[00:51:30] Okay, so three minutes. What is the systematic review process? I mentioned we determine scope and review questions. You're already most of the way towards review questions. Then we collect all the relevant studies. We do searches like of electronic databases for published literature, organizational repositories, websites. We contact authors that we know produced studies to get them to send us their studies, and we do literature snowballing and all those kinds of things.

[00:52:00] So, very rigorous process to collect all the evidence. It's important. If you're not trying to get all the evidence, then it's not a systematic review. Then we determine whether that evidence is includible or not based on our PICOS, and we assess biases in the evidence base through critical appraisal approach. Then we synthesize the evidence and we communicate the findings, the synthesis being the meta-analysis and other approaches.

So, there's the data collection. I'm going to actually ... No, I'll take a question.

[00:52:30] Speaker 16: Hi, [inaudible 00:52:38], global Knowledge Initiative. What if there isn't an adequate number of resources that we are able to collect and there is a gap in the knowledge that is available? Do we reframe our question or reframe our scope in the beginning?

Hugh Waddington: [00:53:00] It's a really good question. This is partly why we developed the evidence gap map approach. Essentially, evidence gap maps are based on a systematic search. It's not as systematic as a systematic review, but it's a broadly systematic search just to give us a sense of, si there actually sufficient evidence here to answer this question appropriately?

[00:53:30] Now, A, an empty review, a review that doesn't contain much evidence or even no evidence, is a useful exercise because it tells us we need more evidence in this area. Obviously as a funder, it's not really what you're after and usually we will commission evidence ... A review that will at least go along the causal change. It won't be focused on a particular outcome so we can maximize our reach.

[00:54:00] So, we collect the studies, systematically published and unpublished literature. We assess bias in the studies to account for other possible reasons why changes happen, problems in data collection, and things like public bias we've talked about. We do not base quality assessment on whether studies are published in peer review journals or not, because peer review journals are biased towards publishing studies with positive findings, and we want to synthesize the entire evidence base, not just that which was able to peek the interest of a journal editor.

[00:54:30] We report that ... Do I wanna take a question? I don't think I have time. Is it urgent?

And we report this information transparently. This is what results look like before and after we control for bias. Biased estimates generally give us bigger impacts. Systematically when we look across studies, this is what we find.

[00:55:00] Now, I was gonna talk to you about synthesis. I was gonna indoctrinate you on meta-analysis, and I'm not sure that we really have time for that. What I would like to say is that vote counting is just wrong. Basically, saying does a study have a positive or a negative or a non-statistically significant effect, adding up that, tallying it and saying the one with the most votes is the intervention that works is just wrong because it just doesn't take into account the magnitude of the impacts or the precision of the impacts.

[00:55:30] It's just wrong. Just don't do it. Do not accept. If somebody comes to you, a research team that gives you a vote count, you tell them to go away and do he meta-analysis. Why is meta-analysis actually policy relevant? It's not just technically right, it policy relevant. Look at this. This is another forest plot. Lots of individual studies.

[00:56:00] You can see that there is this vertical line here. This is the line of no effect. If in individual studies, 95% confident intervals overlaps this line of no effects, then essentially the policy conclusion is, no significant impact. The problem is, all of these individual studies are underpowered. Nine studies here, which were looking it, the impacts of self-help groups from Oxfam, this is Oxfam meta-analysis.

[00:56:30] Nine of them all found statistically insignificant impacts. And if they hadn't done the meta-analysis, they would've included that there's no impact. They do the meta-analysis, which is, they get their diamond at the bottom, and because of the power of all of this additional analysis, they conclude that actually, there is a significant effect. Now that is something which is highly policy relevant. Doing the meta-analysis means you are more likely to conclude that programs have effects.

[00:57:00] Okay, not gonna cover this. The final thing on stakeholder engagement, the approach that we prefer is ... We don't jus disseminate, we don't just produce these plain language summaries at the end. We actually work with you from the start of the project to define the question, to own the evidence. We really want people to use us to own the evidence, and we do that through stakeholder engagement groups. We have specified approaches, which have been proven to be very effective. We have a number of reviews that have use this approach very effectively in getting impact.

[00:57:30] Now we're on the final bit, Anna.

Anna Heard:  
[00:58:00] What we had thought would be fun a little bit, we have placed plain language summaries on all of the chairs. Some of you have also picked up a plain language summary form outside and brought it in. What we wanted for you to do is to pick one, read it, and find other people who also have the same plain language summary for a quick discussion of what the kep points that you took away from this plain language summary. And if there is anything that you felt was really missing

[00:58:30] from the plain language summary that you really still had questions about and you didn't see the answers in the summary at all.

So, if you are a policy maker and you are reading this plain language summary, would it provide you with the information that you are looking for? We're gonna give you five minutes to read your summary, and then we're gonna figure out ... So, who has-

[00:59:00]

Hugh  
Waddington: No, no. We'll do mixed groups. Just do it by ... They can be a group. They can be a group.

Anna Heard: And they just pick one summary?

Hugh  
Waddington: No, no, no, they will have different summaries.

Anna Heard: And then talk about what was good and what was bad about the summary. Okay, so let's make three groups I think. 1, 2, 3, 4, 5 in this group. 1, 2, 3, 4, 5 in that group. 1, 2, 3, 4, 5, 6 in this group. Just read your summary and the have a quick discussion about what you thought was good about the summary, your key takeaways, and what you thought you were still left wondering.

[00:59:30]

Okay, I think it's time to wrap up and it's time for the next session. But what I was hearing and quickly chime in. What I was hearing is that some of these plain language summaries might have been a little bit improved if there were more specific example, or a box that included a couple of specific examples of the intervention that was talked about.

[01:00:00]

One person said something about if they had graphs or charts that visualized the results, that in a quick way it would be a little bit faster and easier to comprehend quickly than the narrative explanation of what they found in the systematic review. Did anybody else have some big takeaways that you either thought was really good or a little bit difficult about these plain language summaries?

[01:00:30]

Speaker 17: Well, [inaudible 01:00:29] them in the sense that the reason they're plain language summaries, we're assuming that the policy makers don't have the time to read very much, but that all of us who read them wanna know more. I think they're kind of teasers in that sense. They're invitations to go deep, or assign somebody else to go deep.

Anna Heard: All right, thank you very much. I don't think we have anymore time. If you have any further questions, you can contact either Hugh or at [hwaddington@3ieimpact.org](mailto:hwaddington@3ieimpact.org).  
[01:01:00] Okay? Great, thank you.