

# Looking Back to Give Better: How GiveWell Evaluates Its Grantmaking

## GiveWell Webinar, June 9, 2026

*This transcript was automatically generated using software, with additional editing afterward, and may contain minor inaccuracies. If you have questions about any part of this transcript, please review the original video recording that was posted along with these notes.*

Elie Hassenfeld: Hey everyone, I'm Elie Hassenfeld, I'm the co-founder and CEO of GiveWell. Just really appreciate all of you joining us today. Thank you for coming to this webinar. We're excited to go through this with you today. Please know that we're going to record this session, and we're going to share it via our email newsletter later this month. So if you'd like to rewatch it or share, you'll be able to do it then. So what are we talking about today? We do a lot of work at GiveWell before we direct funding to an organization because we want to understand how funds will be used and how much impact the program will have. In addition to that pre-recommendation work, we also follow up on how the organization has done after we direct funds to it. And understanding that past funding is a really critical part of the way in which we learn from our own giving, and GiveWell as an organization improves over time. When [00:01:00] we follow up after a grant, we can learn about things that we might not have seen. So we might, we might see that a program ultimately costs more or less per person reached relative to what we expected. We might also find that a program was able or wasn't able to deliver its program as planned. GiveWell has not always done these sorts of backwards-looking assessments of our work as well as we ideally would have, but in the past year, with the growth of our team, we've been able to invest more time and energy into these. And we call these backwards looking assessments of our grantmaking "lookbacks." So today we're going to go through a few examples of our assessments of past grants, what we've learned from them, and how they inform our work going forward.

We're going to discuss four of these lookbacks focused on New Incentives, on Evidence Action's Dispensers for Safe Water program, on the Against Malaria Foundation, and on Results for Development's pneumonia treatment program. And I'll explain a little bit more [00:02:00] about each of these programs as we go. As far as I know, these kinds of systematic post-grant assessment, where a funder compares actual to expected results and then formally applies them to future work, is extremely rare in the grant-making world. And we at GiveWell have learned a lot from these—the ones we're going to go through and also other lookbacks we've done. Some of the things that we've learned have

already changed the way in which we work. And so I'm excited to share this with all of you today.

Here's the structure of what we're going to go through today. In a minute, I'll introduce the panelists and we'll kick off the discussion as we go. Please add questions you have to the chat throughout our discussion. So you know, they'll only be visible to me and to our panelists for your privacy. Also, thank you to those who submitted questions in advance. But if you submit these throughout the session, we'll be able to take as many questions as we can in the last 20 minutes, and we'd really appreciate hearing from you about what's on your mind. So, Adam, Alex, [00:03:00] and Julie. Welcome. You're going to come on, and we'd love it if you could each introduce yourself and your role at GiveWell. Adam. You're up first.

Adam Salisbury: Hi, everyone. I'm Adam Salisbury. I lead our livelihoods portfolio at GiveWell, researching ways to cost-effectively improve living standards of the global poor. Before that, I worked with Alex's cross-cutting team, and I led lookbacks into five programs, one of which we're going to talk about today, which is the Results for Development pneumonia program.

Alex Cohen: Hi, my name is Alex Cohen. I'm a program director at GiveWell. I've been here about seven years. I lead our cross-cutting team. So our team looks for ways to improve our research by critiquing our work or figuring out ways to improve the data we collect to assess our grants. And lookbacks are one of the tools that we use to do this, and so my team has been leading initial lookbacks.

Julie Faller: And [00:04:00] I'm Julie Faller. I am also a program director of grantmaking here at GiveWell. And I work on issues, strategic issues that span across our grantmaking and have been at GiveWell for six wonderful years.

Elie Hassenfeld: Well, thanks Adam, Alex, and Julie. We're going to start with a lookback that we did on the organization New Incentives. And so in brief, New Incentives offers conditional cash transfers. So small cash incentives to encourage caregivers to bring their children to get routine childhood immunizations in Nigeria, in northern Nigeria, which is one of the areas of the world with the lowest rates of routine childhood immunization. New Incentives has been a Top Charity of GiveWell's since 2020, and we, as part of this lookback, evaluated funding that we gave them in 2020. We need some time to pass after the funding goes through for the program

to run and then get the data and understand what's happened. So Alex, [00:05:00] can you, let's start at the beginning with this program. You know, what did we learn from our lookback? What was the core takeaway?

Alex Cohen: Yeah. So the core takeaway for this lookback, looking back on our initial grant to New Incentives was, these grants were about twice as cost-effective as we initially thought. So we predicted that our initial \$17 million grant would avert around 4,000 deaths. So at the time of the grant, that's what we thought. But looking back, knowing what we know now, we think that it actually averted around 8,000. So about twice as many deaths averted. The main driver here was lower costs. So we initially estimated the program was about \$40 for each child enrolled in this conditional cash transfer program, but it ended up being about half that.

Elie Hassenfeld: Okay, so it ended up roughly averting twice as many deaths [00:06:00] as we expected. And that was because it cost half as much as we expected. You know, why did it cost so much less than we expected?

Alex Cohen: Yeah. So we think there were a combination of things here. So one is economies of scale that were bigger than we thought. So as the program scaled, fixed costs like central office staff got spread out over more people. There was some efficiency efforts by New Incentives. So looking for ways to do the program, you know, more cheaply. And then also during this time period, around 2020 and a couple of years after, there was devaluation of the naira, which is the currency in Nigeria. So US dollars from GiveWell's donors buy more in terms of conditional cash transfers or pay for workers that are working on this program.

Elie Hassenfeld: Yeah. So like, what do you make of those findings from this lookback? It ended up being, [00:07:00] I guess twice as cost-effective, meaning twice as much impact per dollar spent as we expected. You know, and as you think about GiveWell's research, what does that kind of finding mean to you going forward? How do you take the specific results from this lookback and think about what it could mean to how we do our research work in general?

Alex Cohen: Yeah, I think the big takeaway for me is that these cost-effectiveness estimates that we have when we're deciding to make a grant are pretty noisy. So I think, you know, there's a temptation when you're at GiveWell and we're deciding whether a program is above our cost-effectiveness

bar, and you know, if our bar is 10x and we do our analysis and we decide, well, this program is 8x, well, that's below the bar, we probably shouldn't make this grant. You know, I think it's easy to see that 8x is like a very precise number and not realize there are these huge error bands around it. And I think this example goes to show that we were off [00:08:00] by 2x in this case. And I kind of remember the early days when we were first considering this grant to New Incentives. It was kind of hovering around the bar and we were like, oh, should we make this grant? Should we not? And, yeah, I think looking back, it would have been a mistake if we hadn't. And I think, I don't know, good to keep that in mind going forward. We, you know, get this cost-effectiveness estimate at the end of our analysis. But it is a noisy number.

Elie Hassenfeld: Yeah. And so I think like something that people sometimes think of GiveWell, because we spend so much time trying to quantify the impacts of programs, I think there have been, we've sometimes been referred to in the media as the the "spreadsheet approach to giving," which I think granted is fair because we use a lot of spreadsheets in our work. But there's also a way in which we recognize the limitations of the quantification that we do, that the quantification is undoubtedly a critical component of our work, arguably the most critical component of our work, but also [00:09:00] has, you know, massive uncertainty around it, as evidenced in this case with New Incentives where the results were significantly more positive than we anticipated. I wonder what you do with that finding. So, you know, in most of the work that you lead, you know, quantification is critical and we're making decisions based on the outcomes of analysis where, you know, you might know that there's a significant amount of uncertainty in your analysis, but then, how do you, what does that do to the decisions you make? And how do you make different decisions given that uncertainty?

Alex Cohen: Yeah, yeah. So I think there are a few takeaways from this. You know, one is that, yeah, we need to think, we need to look for ways to get information that's not included in our model. So yeah, we've got these nice spreadsheets and, yeah, I've definitely done a lot of spreadsheet work in my time at GiveWell that tell us, you know, if we link [00:10:00] up this evidence, this is where cost-effectiveness lands. But can we bring in more qualitative information? So talking more to recipients of a conditional cash transfer program, learning more about the implementation challenges of this program, learning more about what people in Nigeria think about conditional cash transfers—do they view this as impactful? That sort of information, I think, provides a nice

complement to what we see in the spreadsheet. And I think there are also other ways to triangulate our estimates. So one thing that we're working on now is, you know, our models are based on stringing together evidence from RCTs and other sources. But can we also try and measure, you know, real-world today when we're delivering these programs in northern Nigeria, can we detect an effect on mortality rates or vaccine-preventable diseases that would provide like a nice sense check on what we're seeing in [00:11:00] the model? So I think that's one approach is just bringing these other sources of information. I guess the other is like, yeah, it's important to try to learn over time. You know, we're doing these lookbacks to learn. Okay, were we wrong about this? And can we build in more ways to add that learning check? Yeah. Where are we overoptimistic? Were we too pessimistic? Because, yeah, we don't know these things at the outset perfectly. It's good to try and learn over time.

Elie Hassenfeld: Yeah, yeah. And I think as we've, you know, grown the size and capacity of the GiveWell research team, we've been able to do more of these lookbacks. But then also we're in a position where our larger team has the ability to take on some of the projects that can help us triangulate, as much as we can get to the truth—you know, one of GiveWell's core values is truth seeking, just trying to get the best answers we can. And with this larger team, we've been able to go further in gaining [00:12:00] more information to understand, you know, what we expect to take place in the world and what actually has as a result of our funding.

Alex, I think it would be helpful to just talk for a minute about what it actually takes to do a lookback, like what were you able to evaluate, post hoc and what were you not. Because, you know, I do think that this kind of assessment after the fact, this formal reassessment is relatively rare. And I think one of the reasons it is relatively rare and one of the reasons it took us a while to do this at the level of, I don't know, systemization that we're doing it now or consistency that we're doing it now is it's very challenging to gather information after the fact and determine what it actually means. So maybe you could just talk a little bit about that with respect to New Incentives. You know, what were we able to get? What are we not able to get? And what are the big challenges in coming to conclusions about how a grant went?

Alex Cohen: Yeah. So I think about this in like a few different layers. [00:13:00] So we made a grant to New Incentives for conditional cash transfers. We want to know how it went. You know, what we can do is look at the monitoring

and evaluation data that New Incentives collects that measures the number of kids that enroll in the program and whether they stick with the vaccination schedule. We can look at our own research and say, did we learn something over time about the impact of vaccinations in northern Nigeria? We can talk to people on the ground about how the program went. Was there fraud? Were there issues of dual enrollments, that sort of thing? That's kind of the first layer and that's more what we focused on with the New Incentives lookback. But there are a lot of things that doesn't tell us. One big piece is, yeah, [00:14:00] did this program actually lead to reductions in mortality? We've got, again, pieces of evidence that we think link up to that, but that's, you know, we don't have a brand new RCT of New Incentives' program. We've got an older one that we're extrapolating from. So we've, we're missing that sort of thing.

And then I think especially in this New Incentives lookback, we didn't go as deep on getting that on-the-ground qualitative information. In future lookbacks, we've done more of this of like talking to the program implementers, talking to the recipients of the program, talking to local officials, government officials about this program. We've certainly done some of that through the course of investigating grants to New Incentives, but less as part of the lookback. And I think that was one thing that we learned is, you know, this was very much a first iteration of the lookback. And I think we learned through that on like [00:15:00] ways we could go even deeper in subsequent lookbacks.

Elie Hassenfeld: Yeah. Well, great. Thank you so much, Alex, for talking us through New Incentives. We're going to move on to our second case study, looking at Evidence Action's Dispensers for Safe Water program, and where New Incentives ultimately had significantly more impact than we had anticipated at the start, this Dispensers for Safe Water program had substantially less impact than we anticipated when we made the grant. The Dispensers for Safe Water program is a water chlorination program which installs chlorine dispensers near water points that people tend to use in low-income countries to collect water. I personally have had the chance to visit this program in parts of Kenya and in Malawi, and we did a lookback on this program earlier this year. And, you know, to Julie, you'll talk us through some of what we learned and how we were able [00:16:00] to learn more about this program's effect.

And just want to encourage everyone on the webinar, keep sharing questions. Thanks to you who shared questions so far; we have more

than 200 of you here now—donors, grantee partners, and others. And so just really appreciate your interest. Please do keep sharing those questions. So, Julie, why don't we start with just, you know, what did we learn? What was the core takeaway from our Dispensers for Safe Water lookback?

Julie Faller: Yeah. So the core takeaway, as you alluded to, Elie, was that the program was substantially less cost-effective than we thought at the time we made the grant, and that was because although the dispensers were installed sort of like in the numbers that we expected, it turned out that just far fewer people were using the dispensers to chlorinate their water and get safe water than we anticipated. So we now think about a third of the people that we initially estimated were actually chlorinating their water.

Elie Hassenfeld: Yeah. So to start, do you have any idea why [00:17:00] fewer people were using the dispensers than we thought? So we'll talk about how we came to that, how we learned that fact that it was substantially fewer people. But like, why would that be the case?

Julie Faller: Yeah. So we have hypotheses, but I don't think we have a very satisfying answer right now. And I think a lot of the hypotheses kind of revolve around the fact that it's a program that relies on behavior change. So it relies on promoters to promote chlorine, depending on the model, also to fill up the dispensers with chlorination. And then it asks people to consistently use the dispensers and keep their, so that they get the benefits of the treatment. And that amount of behavior change might just be difficult to sustain over a long period of time.

Elie Hassenfeld: Right. And so it's a program where often, you know, people will collect water in these large jerry cans. They'll carry them to the water point, put water into the jerry cans, and there'll be a chlorine dispenser nearby, within arm's reach of the water point. And people have to turn a [00:18:00] dial to get chlorine into the jerry can every single time. Even when we supported this program, I think our baseline assumption was about half the people would be using this chlorine dispenser, meaning we knew that it was difficult to have the behavior change. And then the surprising result was that it was closer to, you know, one third of half is what, you know, so a sixth of the people actually used it roughly compared to the half that we expected. But we certainly recognized the challenge of this, you know, consistent behavior change of the person picking up the water, also the need for chlorine to be present so that it's available to be put in the water. So how did we learn this? You know,

what was the process by which—you know, Evidence Action conducts a lot of its own ongoing monitoring, you know, but how did we get this data in this case?

Julie Faller: Yeah. So it was sort of an interesting series of events. So Evidence Action was collecting their routine monitoring. And we had looked into the protocol under which that evidence would be collected [00:19:00] and sort of thought we understood how to interpret that evidence. Separately, we funded another study in Kenya that overlapped with some of the areas where Evidence Action's dispensers program was operating. And that separate study found chlorine use that was just significantly lower than Evidence Action's routine monitoring suggested. Of course it wasn't a super clean comparison. The village overlap wasn't perfect. Some of the data was collected during Covid, where there were disruptions, and there were other reasons that it was difficult too—different methodologies. So it wasn't as easy as saying like, oh gosh, we really know there's a problem here. There were ways that there might not have been a problem, but that was enough to throw up a red flag and make us want to look deeper into it. So following that, both Evidence Action implemented their own independent data collection exercise to compare that to [00:20:00] the routine monitoring data in Kenya. And we funded a totally separate evaluation exercise where an external group was implementing the same protocol in the same places as Evidence Action's teams, but doing it completely independently.

Elie Hassenfeld: Yeah. And so, you know, what did we take away from that analysis—and presumably the Evidence Action's internal monitoring didn't match up with the independent monitoring. And so why was that the case as far as like, what are your hypotheses or conclusions about why that was the case?

Julie Faller: Yeah. So I think we feel fairly convinced that most of the discrepancy was driven by departures from the protocol as written down. So there are a series of steps that were designed to, you know, make sure that people collect data that's representative. And we think that a lot of the departures might have come around either sampling, like deciding [00:21:00] which households to talk to, or around interpretations of the results for whether chlorine is present in the water source. And there are a few other smaller things, but basically 3 to 4 biases, you multiply them all through and you end up with a pretty substantial discrepancy.

Elie Hassenfeld: And so in your role, you lead or you manage like a number of grant-making teams. And so, I mean, what do you, when you look at the results from this Dispensers for Safe Water lookback, like what, how does that influence your thinking about the grants you're thinking about today?

Julie Faller: Yeah, I mean, I think the very meta framing is that it really matters what people are doing and not just what is written down in terms of what they're supposed to be doing. And so I think that matters both programmatically and when you're thinking about monitoring and evaluation. You know, you can have a beautiful plan, and then human beings, being what we are, like [00:22:00] might do something that's different than the plan. And it's just really important to keep looking for different sources of evidence about what's happening on the ground.

Elie Hassenfeld: Yeah. And I think this is just a good example of a case where, you know, I'm very happy by the extent to which GiveWell has been able to improve our capacity and capabilities to gather and analyze this kind of data. When GiveWell started, we would look at the internal monitoring of an organization like Evidence Action, and we would have seen it as—and it truly is even today, like very strong relative to the vast majority of information that is available and is out there. And, you know, now we're in this position where we can go further and gather more of this information to be in a better position to follow up on these tricky little humans who aren't following, always following the rules of what, you know, what we or others would like them to do. So thanks, Julie, for walking us through the dispensers case study. [00:23:00] We're going to just keep moving through these because we have a lot of ground to cover. And then we'll take questions. We have two more. So we're going to talk again.

Now we're going to talk about the Against Malaria Foundation and a malaria net distribution that it did in the Democratic Republic of Congo. So the Against Malaria Foundation has been a GiveWell Top Charity for a very long time. And it funds large-scale malaria net distributions in areas of sub-Saharan Africa where there is high levels of malaria, and insufficient funding to deliver the nets that people need to cover their sleeping spaces to protect against the mosquitoes that transmit malaria. And so we've given a significant amount of funding to the Against Malaria Foundation in Democratic Republic of Congo specifically because it has a very large portion of the global malaria burden and insufficient funding from other actors to fully cover the need for malaria

nets in DRC. But recently we conducted this large scale lookback of [00:24:00] a grant we made to the Against Malaria Foundation in 2023. And Alex, let's talk through this one. You know, what did we learn from our reassessment and our lookback at AMF's work in this campaign?

Alex Cohen: Yes. So the bottom line here was that we think this grant is about as cost-effective or probably even a little bit more cost-effective than we initially expected. And here we'd expect, we'd estimated something like a 25% reduction in child mortality as a result of nets campaigns. And that was very close to what we found in this lookback. And this one was a little bit unique in that we got to do some of this additional triangulation that we've talked about, where we try to measure the effect of nets campaigns going on in DRC around the time of this grant [00:25:00] on actual malaria rates or all-cause mortality. So that we're not just relying on, you know, the older data that's in our CEA, but actually trying to pull in some live real-world nets campaign data.

Elie Hassenfeld: Yeah. And so that analysis was new and something we hadn't done before. Maybe just take a minute and talk a little bit about how, I don't know, like how we did that, why we hadn't done it before and what was challenging, why we could do it now and what we were able to learn.

Alex Cohen: Yeah. So the basic idea for this analysis was to use staggered timing of campaigns across provinces in the DRC as kind of a natural experiment. So we're comparing mortality outcomes for children born just before versus just after nets campaigns come to their province. This is, you know, this is the type of analysis that we [00:26:00] talked about doing before. This is something that was available to us before. We had this source of variation. We had the DHS data. The DHS is a large-scale survey of health outcomes that happens across countries. That's what we use to measure malaria rates and mortality across provinces. These all existed before. But I think the big difference has just been staff capacity. You know, we haven't had as much bandwidth to devote to this sort of lookback or checks of our work. And second, we got a lot of use out of AI for this. We used Claude Code to help with the analysis—certainly not totally unsupervised. There's a lot of checking that needed to happen, but I think that also helped make this work a little bit more feasible in a way that it hadn't been before.

Elie Hassenfeld: And I think one of the things that came out of this lookback that was most salient to me is that [00:27:00], you know, being born in DRC, you know, right after a malaria net campaign means that you'll have

significantly higher chances of survival in that first year, which is a critical year of life relative to being born in the year before the malaria campaign. And it was just a really compelling reminder of how critical these programs are to, you know, to help people and how substantial of an impact it can have on people's lives where, quite literally, people are living and dying based on the year in which they are born and whether they have access to these needed health tools. You know, another part of this lookback that I thought was really interesting was the work that we did to get qualitative information as a way of checking this. So, you know, we did a lot of quantitative work. We've done a huge amount of technical work on malaria over the last, you know, 15 plus years. And then we did this additional quantitative work in [00:28:00] this case, as you described. But then we also did a fair amount of qualitative work here to try and understand what we might be missing in the quantification. And maybe you could just share more of that.

Alex Cohen:

Yeah. So that was the other spot where we went deeper in this lookback than we have in other lookbacks, which is qualitative information. So practically, we worked with a survey firm based in DRC, sent a team of surveyors to, I believe, five provinces in DRC to go and talk to people about net campaigns. And we prioritized in particular places where there had been challenges with net campaigns. We wanted to learn, you know, what could go wrong here. And so we heard lots of stories here, stories of nets being used to create log rafts or some instances of fraud or some instances of favoritism in the distribution. You know, if you know the distributor, [00:29:00] you might get more nets than if you don't know the distributor. And I think, you know, this is something that we hadn't done before in large part because, you know, this requires also processing a bunch of surveys and transcripts. And that's something we can do now with AI that we couldn't do before.

But in terms of the takeaways from this work, I think, yeah, it was kind of good confirmation that the world is complicated, especially when you're working in a country like DRC. Challenges come up. And I think we also heard cases of people overcoming those—getting in a boat to get out to people that couldn't be reached by a motorbike. But then I think it's also a good reminder that even, you know, with these issues, we were seeing pretty credible evidence of nets being delivered, effects on burden and mortality. And I think, [00:30:00] yeah, it's possible for both things to be true.

Elie Hassenfeld:

Yeah. And I think sort of, in aggregate, you know, we tried to support this additional survey work to really find evidence of problems and get a better feel for the level of problems. When we do our grantmaking, we assume there will be some significant degree of problems. Certainly not every net that is distributed will be hung over a sleeping space and be utilized consistently. And that's, you know, built into our assessment. And the programs are extremely impactful and cost-effective, taking into account that loss. And so I think, roughly speaking, we saw the magnitude of identified problems via the qualitative work as being consistent with our preexisting expectations of what would happen in this case. And, you know, for me, in some ways, it's very confidence increasing, from my perspective, in the grant making that we've done because we go out and see evidence of problems. And I think if we were finding [00:31:00] and when we have in the past sometimes asked grantee partners for evidence of problems, when we hear there are zero problems, everything works perfectly. That's not how the world works. And so we know that that's not, you know, that's not quite right. But I think this is just another example where, you know, for a long time, GiveWell has tried to be at the vanguard of the level of rigor that we're bringing to our work. And I think in the last few years, as we've expanded the team and the scope of what we do, we've been able to take on even more.

So thank you, Alex, for, for taking us through the AMF case study here. I want to turn to our final case study before we then turn to questions, which is focusing on Results for Development and its pneumonia treatment program. And this program is a little bit different than many of the programs that GiveWell tends to support and GiveWell is most known for. It's a kind of program that is sometimes referred to as technical assistance. And it's called technical assistance—that's, I guess, a term of art in the global health world—because instead of the [00:32:00] charitable dollars being used to directly purchase and then deliver the health good—so in this case it would be delivering amoxicillin through health facilities—instead the charitable dollars are used to support the government in its ongoing delivery of the needed health good through the health system. And we have done a substantial amount of grantmaking in this space of technical assistance, supporting government delivery of health. And so we decided to talk through this case study in Results for Development. There are other case studies of technical assistance work that we've done that you can see on our website. But Adam, why don't you join us, And we would love you to just

explain what we learned from looking back at this Results for Development pneumonia treatment grant that we made.

Adam Salisbury: So I think the big-picture takeaway was that this grant broadly met our objectives. And we think it was about as cost-effective as when we modeled it in the first instance. [00:33:00] The main thing, the main objective of this grant was to increase the supply of amoxicillin in Tanzanian health facilities, and we think that was accomplished. So, supply went from about 50% in 2017, when we first made the grant, to about 90% in 2022, when we last funded a survey. Us and Results for Development have since been keeping tabs on amoxicillin supply in Tanzania, and supply seems to be hovering around 80 to 90%, which is great. The big caveat with these types of technical assistance grants is it's really hard to know whether GiveWell and R4D was causally responsible for that increase, or whether the government would have increased the supply anyway. We tried to get at this in creative ways through this grant, which I can get into. But the bottom line is that even when you think about it carefully, it can still be really hard to come to confident conclusions.

Elie Hassenfeld: Yeah. And so let's talk about that a little bit because in the case of say, the Against Malaria Foundation, you know, there's very strong reason to believe [00:34:00] that a place like DRC has insufficient funding for nets, or there are states in Nigeria that have not had net campaigns for a very long time. And therefore you say, well, the funding that we're providing is causally increasing the availability of this needed health good when we're directly supporting the supply of something like a malaria net. And then there are randomized controlled trials—in the case of malaria, many, many randomized controlled trials—showing the effect of what having those nets does to malaria cases and mortality. But then how do you think about the question of causal impact and, you know, of our support in a case like this technical assistance grant for pneumonia?

Adam Salisbury: Sure. I mean, so the fundamental question we grapple with is what would have happened otherwise. And what's tricky about these grants is that the government actively wants to increase the coverage of these programs. So in this case, they came to R4D, and who came to GiveWell and said, we want to increase the supply of amoxicillin in health facilities. Can you help us do it? Given [00:35:00] they want to do this anyway, the question we ask ourselves is like, did R4D make this happen, or would this have happened anyway? One way we tried to probe at this for this grant was to monitor the coverage and availability of other medicines

that weren't targeted by this grant as a proxy for what would have happened to amoxicillin had R4D not provided technical assistance. And so we tracked the coverage of paracetamol, of benzylpenicillin, and of zinc and oral rehydration salts. Generally speaking, the coverage of these commodities went up as well between 2017 and 2022, but they increased at a slower rate than the supply of amoxicillin increased, which we take of as some evidence that R4D's work was causally responsible for increasing the supply of amoxicillin faster than it would have increased anyway.

Elie Hassenfeld: Yeah. And so then what do you make of this with respect to the technical assistance grant? You know, the first three case studies that [00:36:00] we talked about, while there's some degree of uncertainty about what the causal effect of our funding was, we have, you know, more of a hand, something to hold on to, something to grasp that shows this really had an effect. And in this grant, you know, we're, you know, we have our best attempt, which is comparing amoxicillin to these other medicines. Like, how do you think of this in general and what leads you to think either—I mean, would you support a grant like this again? And if so, why with that level of uncertainty?

Adam Salisbury: Yeah. I mean, I think I would support a grant like this again. I think for these type of grants, what's really important is really interrogating the theory of change, so really trying to understand what are the bottlenecks that prevent governments scaling this up as quickly as they would with TA, and how did the technical assistance activities map to these bottlenecks to scale? And I can give you an example of a story that we found in this case. So I went to visit Tanzania about 18 months ago. I spoke to both R4D and [00:37:00] I spoke to officials in the Tanzanian Ministry of Health, and I asked about what exactly did R4D do that helped you increase the supply of this drug? One example they gave is that they helped the government improve their forecasting of the demand for amoxicillin. So if you think about when the minister is deciding how much amoxicillin to buy, that's a function of the demand for amoxicillin, which is a function of how many people live in rural areas. It's a function of care-seeking rates. So how many people go to facilities. It's a function of the prevalence of pneumonia, which varies year to year. So it's a pretty hard thing to estimate. R4D essentially revamped the model that the government was using to predict this demand. Essentially, they thought the government was underpredicting demand, so weren't procuring enough amoxicillin as a result of that. That was independently verified by the government minister. It's [00:38:00] not

sort of RCT gold standard evidence, but that does seem to be pretty credible qualitative evidence to me that there were legitimate bottlenecks that were alleviated here.

Elie Hassenfeld: Yeah. And I think there are two parts of this that are, I think, interesting to pull out. So one is, you know, as we've been talking about throughout this conversation, the quantitative data itself is never full and complete in making the case that something is or is not having impact. And in this case, the quantitative data itself, for example, the estimation of impact and then the comparison of amoxicillin to other drugs, probably does less work than the quantitative does in other cases. And so we're then in a position where the only way to get more information is for you to go to Tanzania and essentially interview people like a journalist to do the best you can, to ask hard questions and understand the impact that the program may or may not have had. And I think that's another case where as we've increased our capacity over the years, we have [00:39:00] more of an ability, you know, for someone like you to be able to go there and do that kind of research that was, you know, sort of in an extremely time-consuming way. Just as a final question to you, Adam, how does this make you think about technical assistance grants in general?

Adam Salisbury: So I did this project as a part of five other lookbacks of technical assistance grants. One big takeaway which came out in this grant as well, is that I think it's important not to overanchor to very specific and linear theories of change. So if we make this grant, then the government will do that and then coverage will increase by this amount. I think government priorities change all the time. So, in this grant, the executive changed through the process of making it. Rather than try to predict exactly what the decision making will look like, I think we should, and I think we are, putting more emphasis on more reliable and more robust predictors of success. So things like the strength of relationship [00:40:00] between the grantee and the government. So like, have they worked with the government before? When we speak to the government, do they recommend them? Can they introduce us to the right people, etc.?

I'll give an interesting example, just to conclude in this case of this grant evolving in directions we just didn't anticipate at the start. So when we went to Tanzania and interviewed both R4D and government officials, they said that one of the main benefits they'd got out of this grant was that R4D were able to help them with other demand forecasting tools as well. So not just for amoxicillin, but also for bed nets and for antiretroviral drugs, which they think improved resource allocation there as well. That's

just something we did not expect would happen, would have happened at the start of the grant. And so I think it's, yeah, another reminder to not overanchor on being able to perfectly predict the theory of change.

Elie Hassenfeld: Yeah. Thanks, Adam. And then I think like, one final thing to just pull out about these technical assistance grants is, you know, on one hand, what we take on with this sort of grant is a [00:41:00] weakness, so to speak, in how confident we can be in the causal impact that our funding is having. And then the benefit that we gain is by working with an organization that is trying to support the government, the country, its health system directly, it has more of the opportunity to be, to have impact that goes beyond just the time frame of the grant. And so in some ways, we're taking on some of that additional risk and uncertainty for some additional benefit in trying to do the best we can in assessing it as we go. And I was really excited when you did this set of lookbacks because they looked at some of our earliest technical assistance grants and were able to, you know, evaluate for ourselves, like how we felt about this approach that we've taken. And I think it's been promising enough that we intend to continue it.

With that, I want to move to Q&A from the audience. Thanks everyone for sharing questions. Please continue to do so. We have a lot of great ones. I will direct these questions to the panelists. And then when I direct these to you, if you don't want to answer them and push it to someone else, you know, [00:42:00] please feel free. But I'm going to do my best here. So Julie, the first question is going to be for you. So we shared that both qualitative and quantitative factors influence our grantmaking. They influence our recommendations. I think it's fairly self-evident how quantitative factors drive our decision making. You know, we have a cost-effectiveness model and we estimate the impact per dollar. And that is a major part of our thinking. But how about times when qualitative factors have driven decisions? Can you think of an example or two where that took place?

Julie Faller: Yeah, definitely. Well, so the first thing I would say is I think they're not as easily distinguishable as to say this is a quantitative or qualitative factor. Like what we're trying to do is get to our best estimate of the impact of a grant. And so that means like using qualitative information to try to adjust our quantitative estimates so that we feel like it reflects what we think is actually going to happen. I think this can show up in a bunch of different ways. Some of those are factors [00:43:00] like Adam was just talking about, you know, is the grantee well placed to carry out this work

because it has appropriate relationships or because it has a track record of being able to carry out operationally similar programs. And so we say, like, if the track record is really strong, then maybe we think there's a higher chance of success that shows up in our quantitative model. And then I think, okay, I'll shift from talking about generalities and just give two recent examples. So one recent example, we were speaking with a stakeholder in an area and we were asking about an implementing partner's, how well positioned they were to provide technical assistance to this particular government.

And the stakeholder actually said, you know, like there's a window of opportunity here, but here are some factors about how the kind of relationships are organized in this space, that maybe there are other groups that are better placed to carry this out. So that's an example of something where we learn something that—and we didn't just [00:44:00] take that person's word for it, we asked more questions and spoke to other people. But that's an example of something that can affect our decision making. Another example is we were considering a program recently in an area that had modeled estimates of burden for vaccine-preventable diseases were quite high. They indicated that these diseases were really a problem, but also really good recent evidence indicated that vaccine coverage was actually quite high, that there weren't major problems with the vaccination quality, which sort of makes you think, okay, so this is inconsistent. Kids are getting the vaccines, but vaccine-preventable diseases are a really big problem. What do you make of that? And in that case, we felt the balance of evidence was, indicated that the modeled estimates were probably just too high above what they should be, which affected our decision not to support that particular program.

Elie Hassenfeld: All right. Well, thank you, Julie. Alex, the next question is going to come to you, do. This is a question about large language models—use of AI. So these are tools like [00:45:00] Claude, GPT, Gemini. Are we using LLMs to measure support, to support grant lookbacks and research into potential future grants? So you know, where and how are we using these tools? And you mentioned a couple of examples, but maybe you could go a little bit further too.

Alex Cohen: Yeah. So, at a high level, that's something my team's thinking about, is how can we use AI to do our research quicker, better. Maybe two examples I'll give. One, we in our cost-effectiveness models, we often need to pull in external data sources. One is this DHS data, which

collects indicators like net coverage or chlorination rates across countries. And, you know, it can be challenging, somewhat annoying to go to the DHS page and download the data and find the right indicators that you want to use. And so some folks on [00:46:00] our team built a tool using Claude Code that downloads all that data and puts it into this easy-to-use format that a grantmaker or researcher can query and say, hey, I want to know, you know, the malaria rates in these areas for the past five years, and they can more easily pull that data. So that's a case where we're trying to, you know, save time with these models. And then the other example is using AI to critique our work. So my team does a lot of red-teaming, trying to poke holes in our models. Is there something we're missing on the case for bed nets? And we found that the large language models, Claude, GPT, Gemini, have gotten much better at spotting potential flaws that we should look into. So that's another path that we've been exploring.

Elie Hassenfeld: Yeah. Thanks, Alex. In some ways, I think that the AI tools are one additional way in which we've been able to extend and expand our capacity [00:47:00] in the short term, because the tools enable us to do things that we couldn't do ourselves or couldn't do as quickly in the past.

Adam, I want to direct this next question to you, which is, you know, about doing these lookbacks and I think it fits well with the technical assistance lookbacks that you did, which is, you know, how do we distinguish between different reasons that a program might not have worked out? So it could be organizational inefficiency, you know, the organization didn't do the right things, but it also could be things like external external shocks, like extreme weather or political turmoil. You know, how do we disentangle all the things that happen and what that can mean for a program's success when we're trying to assess the success of a program or a particular grantee in a specific window of time?

Adam Salisbury: Yeah, it's a great question. So when we do these lookbacks, I think of them as a bit like a post mortem. So we look at all the things that happened that we didn't expect at the time of making a grant. And I think it's important [00:48:00] to distinguish what could we have reasonably predicted versus what was pretty unpredictable. This R4D grant is an interesting example. So these grants span 2017 to 2022. The Covid-19 pandemic happened throughout that window that actually affected this grant. So timelines had to get shifted back. Funding that we'd initially earmarked for technical assistance we agreed to pivot to fund drug procurement directly because the government had to redirect

funding to support the Covid-19 effort. We agreed to doing this because we thought the case was reasonably compelling. It also meant the timeline shifted and we had to make an about \$1 million top-up grant. In hindsight, I think it would have been hard to predict something like Covid happening. So that feels more like a category of unpredictable shock. There have been other examples when we've been doing these lookbacks of things that we think we really should have predicted ex-ante. So one of the grants we made was to support the rollout of the malaria vaccine in Kenya. Timelines were also blown in that grant because of [00:49:00] a presidential election in Kenya. That's the type of shock that we think is very predictable. Like we should have seen that coming. We should have anticipated that that probably is going to delay things. So, yeah, I think it's important to distinguish between those.

Elie Hassenfeld:

Yeah. And I think like one of—there's like two parts of this that seem really salient to me. One is, you know, this is another case where we're trying, we're not just, you know, applying, seeing results and taking them as a given. It's a function of the context in which they take place. And that means that as with all of the work that we do, there are significant subjective, you know, there's significant judgment that goes into the decisions that we make. But I think that's necessary to make good decisions. And then we're also learning about things like the elections, which, you know, we should do a better job anticipating to do our work more effectively.

Julie, we're going to come back to you with the next question, which is how do we think about the appropriate timeline for a lookback? You know, on one hand, capturing data during and immediately at the end of the grant can make it easier to connect impacted grant activities. Waiting longer means that maybe some of that data is less available. [00:50:00] But on the other hand, you have better insight into the true effects of the program. And so just maybe talk a little bit about, you know, these lookbacks entail a lot of work, and so how do you think about the right time to take them on?

Julie Faller:

Yeah. So I think what we're trying to understand is, was the impact of the grant the same as what we thought at the time we made the grant. And so that means that we should be thinking about what was the time span of the impact that we had in mind. So for some grants, we are expecting the majority of impact to come during the funded period. It's, you know, direct delivery programs where we really think the value is accrued right there. And for a program like that, a lookback can be, I think, pretty

proximate because you're just sort of like, did the program happen as we thought? And we might be missing some longer-term benefits, but I think that that's sort of not captured in what we immediately, what we thought at the time we made the grant. For other [00:51:00] programs, if we expect that a lot of the benefits come by virtue of causing government to take up new programming areas that it wouldn't have otherwise, I think you would need to wait a bit longer and find out like, did the government, do government health care workers continue to deliver the program? What's it look like in terms of quality and access to that programming area? And then I would just say, like, finally, I think in my job, I think a lot also about not just were we right or wrong about a particular grant, but also what can we learn from it, and how can we make sure that we're updating the way we approach things generally? And so for me, I think that means that there's a lot of value to trying to take back and think systematically about what we got right and wrong on the earlier rather than later side, because that just lets us notice patterns in ways that we might be missing things.

Elie Hassenfeld: Yeah, totally. Okay. The final question that we have for each of you, and I want to take this to each of [00:52:00] you in turn. So maybe Adam, then Alex, then Julie. Is just like looking back at the lookbacks that we've done overall or the lookbacks that you yourself have been involved in, how do you expect that this will influence Givewell's research in the future? So answer that however you'd like, but I'd love to hear from each of you on this, I think in some ways, the most important question when it comes to lookbacks.

Adam Salisbury: So I can go first. So my lookbacks to technical assistance grantmaking are definitely influencing active grant investigations. Now, one of the grants I'm looking at in my new livelihoods role is whether we should support advisers and consultants to help the Indian government scale one of their social welfare programmes. Rather than build a detailed cost-effectiveness model, which I think we should have shifted more towards in the past, I'm spending much more time interrogating the theory of change. That's looking like speaking with government officials, speaking with the grantee, hopefully going out to India later this year to actually trace [00:53:00] the supply chain of what needs to happen to deliver this program. I think that maybe I wouldn't have updated in that direction had I not done this project.

Elie Hassenfeld: Great. Thanks, Adam. Alex, how about you? How are these, how do you think about these influencing our direction in the future?

Alex Cohen:

Yeah, I think I'm going to start with the high level and then get to specifics. So I think the ideal is that the lookbacks let us get better as a grantmaker over time. So, you know, we can see, you know, which things were more cost-effective. We dedicate more funding to those. Which things are less cost-effective. We dedicate less funding to those. We can kind of get better over time as a grantmaker by doing these. But I think one other thing to add here is, I think these lookbacks also point to specific program improvements. So we talked about the nets lookback and one, you know, one takeaway there is this campaign was effective. You know, we should probably continue supporting [00:54:00] nets, but we also can confirm this finding about how, you know, whether you're born before or after the campaign really makes an effect on your coverage. And, you know, coverage of nets declines over time. These are done typically every three years. And by the second or third year, coverage is pretty low. And that really points to an opportunity for us. Should we be looking for ways to provide more frequent campaigns or provide campaigns through a routine distribution like antenatal care. And, you know, these are questions that we're thinking about. We haven't made any decisions yet. But I think the other thing about these lookbacks is it's not just, you know, keep funding or not funding a particular program. I think that's valuable too, but also looking for specific ways we can deploy funding to get more coverage, increase the impact of our grantmaking. Yeah.

Elie Hassenfeld:

Great. Thanks, Alex, Julie.

Julie Faller:

[00:55:00] Yeah, I think I agree with Adam and Alex, and I would just say, another thing that's really on my mind is thinking about whether we have set ourselves up to be able to learn well. There are a lot of decisions that you have to make sort of at the beginning of the grant, thinking about what external monitoring, if any, are you supporting, like what sort of relationships are you going to need to draw on to understand how implementation actually went throughout the course of the grant? Like obviously, what questions are you asking the grantee and data you're asking them to collect. And so I think that's another way where I both feel like the lookbacks are helping us do that better, and I think that this will also help program implementation go better over time, because we'll be sort of like actively monitoring, setting up opportunities for grantees to notice improvements that they can be making over time. Yeah.

Elie Hassenfeld:

Yeah. And for me, I mean, I think that, you know, so much of what GiveWell does is look at organizations and ask what their track record is,

you know, what they have done, as a mechanism for understanding [00:56:00] what they can do in the future. And these lookbacks allow us to turn that same light on ourselves and say, what is our own track record? And I think this is just a critical part of being an organization that is trying to be systematically, assess our own work, be transparent about it so that we can improve over time. I wanted to close this with just one final question that we got, which is, you know, what does success look like for GiveWell in the long run? And, you know, when I think about the place that GiveWell is today, you know what, when we look at the world, we see a world with just a huge amount of need for assistance. You know, to give one figure, there are roughly 4 million annual child deaths in low-income countries that are preventable. You know, we know how to prevent these deaths. We just have not yet marshaled the resources and the knowhow to actually do that. If you think about what 4 million is, that's more than 10,000 [00:57:00] deaths every single day. And this is something that if it were not happening every single day, would itself be front-page news.

And so as GiveWell has grown over time, we're very gratified and thankful for the support that donors like you have provided. Over time, you know, we've been able to make larger and larger inroads against these significant problems. You know, we expect that, well, 2025 was the year in which we raised the most funds ever. We expect that 2026 will be the year in which we direct the most funding that we ever had. And over the long run, our hope is just to make as big of a dent in the problems that exist in low-income countries as we can with our people and the resources that we're responsible for. That's for issues like child deaths, but also the other issues, you know, health problems, income, and other issues that people face in low-income countries. And so we're just hoping to make as big of a dent in those problems as we can with the resources that [00:58:00] we have. Thank you all again for your support of GiveWell and for joining us today, for all the engagement and the questions. We really appreciate it. We don't take it for granted at all because, you know, this kind of wonky take on charitable giving is rare. So thank you all for joining us and for your support of our work. We really appreciate it.