| | |
|---|---|
| Amanda Glassman: | I just read a poll that 64% of parents worldwide believe their children will be worse off as adults than they are today. Disasters, disease outbreak, corruption, economic woes sometimes they seem to conspire to create an atmosphere of pessimism. This pessimism actually has very little place in global health. In fact, the world is getting healthier every year. |
| [00:00:30] | |
| | Let's look at some of the specs. In the past 15 years the annual number of people newly infected by HIV has fallen by about a million people. Deaths from malaria in Africa have been cut in half. We have seen child mortality drop in half. This is the graph that Bill Gates says convinced President Trump to sustain aid. This is truly remarkable, an incredible testament to the work of many people. |
| [00:01:00] | |
| | There's so much more to be done because the questions is, how has this happened? Why have some efforts been so effective and others not? And why is it so important to know? First, because the funding landscape has changed and our aid and public monies are precious. Many developing countries are growing economically, aid is winding down or phasing out, and governments are beginning to spend more of their own money but the demands for the uses of that public money are increasingly intense. People's education is growing, they have more information, there's more urbanization and so these demands are competing with each other for that very scarce public money. |
| [00:01:30] | |
| [00:02:00] | |
| | As aid declines what should we advise a country like Kenya? Build more schools or sustain US funded public health programs like PEPFAR. What can we, as external funders and advisors, really recommend? What is the most cost effective use of those precious tax payer monies? Second, aid obviously is still incredibly vital in the poorest countries. In these countries where the needs are so vast spending wastefully or ineffectively is a tragedy we must do everything to avoid. The truth is that money, whatever its source, does not translate automatically into better health or development for that matter. So much of the spectacular reduction in child mortality is due to economic growth and not only to health interventions and so some people question, very reasonably, does aid work? |
| [00:02:30] | |
| [00:03:00] | This was the impetus for writing Millions Saved. Today, I'm talking about the second volume of Millions Saved but there was a first volume that was authored in 2004 by Ruth Levine and Molly Kinder and a working group of illustrious economists and experts. That book focused on what was working in global health and what aid had to do with it. Ruth and her working group set out five criteria and they selected 17 programs to highlight. They chose programs that were large-scale, that addressed an important public health problem, that had an impact on health, that had lasted at least five years, and used cost-effective technologies. |
| [00:03:30] | |
| [00:04:00] | This edition of the book was pretty successful. It's being used in 60 universities as a textbook that goes along global health programs. It's being used, it's criteria were being used initially by some of the philanthropists that were advising each other really on how to invest in the most cost-effective interventions. It's been a response to aid skeptics. I think aid skeptics number one would be Bill Easterly and |

[00:04:30] he himself was part of this working group so you know it's gotta be rigorous. He sometimes cites Millions Saved when he talks about what effective aid looks like. Then, there have been some follow-on projects, CGD didn't have that much to do with them but [IF-PRE 00:04:31] issued Millions Fed and Brookings, last year issued Millions Learning, that's a big impact for a think tank.

[00:05:00] This new volume builds on that success. We wrote this book a decade later with Miriam Temin and some CGD authors. We took advantage of the much richer evaluation resources we had available. In 1995, which was basically the period through which the previous version was able to look to find evidence, there are about 10 evaluations, 10 impact evaluations that were being produced each year across all sectors. Now, that number is upwards of 300 per year so the evidence base is totally different.

[00:05:30] Second, we wanted to address the strength of the evidence that underlay the programs' achievements, and we wanted to show how rigorous evaluations could be designed cheaply and ethically alongside at scale programs. Sometimes we hear a lot of people say, "Oh, it's just a little NGO project. It's a boutique project. It's not generalizable." None of these programs and their evaluations can be subject to that critique.

[00:06:00] Third, we didn't want to just look at cost-effective technologies because we know a pill works to cure a disease. We wanted to look at the issue of wasn't cost-effective in implementation? Did the delivery work? Then, we added something that people like to talk about but rarely do and that was to look at disappointments. Rigorously evaluated programs that were implemented well but did not have a chance to have an impact on health outcomes. I think those programs are especially interesting because they show exactly why rigorous evaluation matters. What if we do everything right but it doesn't make any difference? We stop doing it.

[00:06:30] Here are the cases. Here's how we selected our cases. We used the same criteria that the first edition of Millions Saved had done but we added a couple of nuances. In terms of important health problems we tried to take big problems that were defined by disability, adjusted life years saved, or other health outcomes. One of the problems is then you don't pick things that have been eliminated. Let's say,
[00:07:00] there's no more measles cases in countries that have high measles vaccine, but that doesn't mean they're not important because they're preventing all of those cases so we made some exceptions in that case.

[00:07:30] We looked for attributable impact and we tried to set a very high bar. We said only programs and policies that could show a positive, statistically significant positive impact, attributable impact on a population health outcome would go in. That was one of the hardest criteria to use actually for a couple of reasons. One, because there still aren't that many impact evaluations of these kinds of programs using rigorous methods. Second, because some programs are universal or some programs want to eliminate all cases of disease and therefore you have only five cases of disease to look at as your outcome. We tried to nuance it a little bit but for
[00:08:00] the most part I'd say about 85% of the cases included use RCT or quasi-experiment

to show their effects.

Third, we took scale and duration. We required more than 80,000 beneficiaries, that's a somewhat random number. We wanted these big populations, we wanted to show that at scale program. One of the troubles with using this criteria is that a lot of pilots, many of the evaluations that are out there, the best of evaluations are based on small populations or small-scale. In some cases, we took programs that had a small-scale pilot that was later scaled and then followed up in a way that we thought was rigorous.

[00:08:30]

We looked at programs that were highly cost-effective and we put a number to it. We said we don't take anything in that costs more than one GDP per capita in that country to prevent a disability adjusted life year. The trouble with that idea is that most of the evaluations and most of the programs have not documented the cost per dally that is required to achieve their outcome. There were only two evaluations of the ones that we looked at that qualified on all these other measures that calculated cost-effectiveness. We had to back out to some estimates. They're not terrific but it gives you a sense of how these programs perform.

[00:09:00]

Finally, we tried to select programs that would be generalizable in the sense of being relevant to other areas, other sectors, other countries. We don't use a formal definition of external validity but we thought they would be relevant. Out of a list of 500 candidates here are the programs that were selected under success at scale and I've highlighted those programs that were US government supported at least in part.

[00:09:30]

We grouped them into four categories, and I'll talk a little bit about why we did that. One category is the traditional medicines and technology rollout. We looked at targeted transfers that can improve health. We looked at behavior change to reduce population risk and we looked at programs that expanded access. Remember, all of them have to change health outcomes. I think this is important because a lot of times we think of global health only as medicine and technology rollout.

[00:10:00]

Here's what we found and this is the list of disappointments, just to say, it wasn't popular to be on this list as you can imagine but I think it's important to show that in every category that we looked at there were many, many successful programs but there were also some disappointments. As I said before, most of them were implemented well they just didn't deliver on the health outcomes that we expected.

[00:10:30]

Here's an example of one of the medicines and technology cases. This is polio elimination in Haiti. This was with support of the US government, a massive vaccination campaign using proven tactics, a very poor country, very difficult circumstances as you can imagine, and basically stopped outbreak in their tracks.

[00:11:00]

In Botswana, we saw ... This is a program that was supported by PEPFAR, the global

| [00:11:30] | fund MERC, and The Bill and Melinda Gates Foundation. It's a mass antiretroviral medicines program and it reduced AIDS-related deaths from a high of about 21,000 in 2002 to less than 6000 in 2013 that's reverting a 15 year drop in life expectancy that happened with the spread of AIDS in Botswana at the start of the outbreak. The program's name [MAS-SA 00:11:55] means new dawn in Tswana. It showed |
|---|---|
| [00:12:00] | really importantly, especially for the history of USAID, that adherence to a complex medicines schedule was possible in a low income country and it led to the scaling of PEPFAR, an important program. |

|  | Technologies don't have to be a drug, they can also be a bed net to prevent against malaria in Zambia, for example, where the US president's malaria initiative has made significant improvements in all cause child mortality using a combination of |
|---|---|
| [00:12:30] | bed nets and other interventions. Or, one of my favorite programs, something as simple as a cement floor to prevent kids from playing in the dirt. This program was supported by the government of Mexico and it not only reduced parasites and iron deficiency among participating children but it also enhanced their mothers' self-reported mental health. It inspired a new social enterprise that's supported by |
| [00:13:00] | USAID DIV, Earth Enable, which is doing improved flooring in Rwanda. That's how evidence can be used in an interesting way to make a difference at scale. |

|  | These are the programs ... This is an example of a program in the category of increasing people's access to healthcare including the universal health coverage agenda. These programs feature expanded use of interventions like they incentivized vaccinations and attended births, they reduced out-of-pocket spending |
|---|---|
| [00:13:30] | on the part of the poor, and they markedly improved health itself. It's not always the case that insurance or subsidized access to healthcare will improve utilization and outcomes, these programs did. Argentina, Rwanda, Thailand. This program in Argentina reduced neonatal mortality by about 20% and that's starting from a baseline that looks like some US states and cities and that effect size is enormous. |
| [00:14:00] | What can we learn also from these very ridiculously evaluated programs elsewhere? |

|  | Because health is a linked to so many social determinants from income to education efforts outside the health sector make a real difference in improving people's health. The most notable of these strategies, and maybe you saw the video that started just before we came in, is cash transfer programs. I started my |
|---|---|
| [00:14:30] | career working on cash transfer programs in Latin America. It was just one program in Mexico and through rigorous evaluation in many, many settings these programs are now all over the world. Millions Saved talks about a program in Kenya. |

|  | As the video mentioned, 480,000 people are benefiting now. It's a small stipend, $20 monthly given to the guardian of children who were orphaned during the AIDS |
|---|---|
| [00:15:00] | crisis. The cash allowed these kids to be fed and go to school. It reduced the odds of sexual debut among adolescents by 30%, which is a huge risk factor for HIV/AIDS. Young women were less likely to get pregnant and young men saw improved mental health. It's huge and it's run by the government itself. In a place that people say is corrupt we're seeing incredible results supported by UNICEF and |
| [00:15:30] | the World Bank. This man's name is Linus, by the way. |

| | |
|---|---|
| [00:16:00] | Finally, policymakers are increasingly recognizing the importance of behavior change for longer lives and better health. This is rough, we know changing behavior is hard but there are several examples of what has worked at scale. Indonesia's sanitation program, which I'll say something about, to Thailand's tobacco tax, Vietnam's motorcycle helmet law, all of these things show that policy changes can be just as powerful as a program that we set up and administer in terms of health. |
| [00:16:30]<br><br>[00:17:00] | Indonesia's program, which was supported by the World Bank and the Bill and Melinda Gates Foundation was one of a series of programs that combined latrine building plus social marketing to reduce diarrheal diseases and related conditions. They were all tested at scale in different countries. I think this is a really interesting example of a positive set of evaluations intending to learn across context what works. The results are impressive. 38 million people benefited. It was extremely cheap, $14 million and they found a 30% drop in diarrheal prevalence. This was a controversial program to include because in its social marketing strategies that it used they tried to shame people basically from public defecation. You can see this poster that was used. It was effective but it's certainly something where you think, "Hmm." |
| [00:17:30] | Knowing that and having the qualitative evaluation alongside the quantitative impact evaluation allows us to consider really how is this program working and how can we do no harm first while also achieving our health goals? I really like what Louis Boorstin, then the WASH Director at the Gates Foundation said about this evaluation initiative attached to the program. He said, "The concept of scaling up, expanding successful small-scale programs to reach entire populations is a myth. We did lots of pilot programs in the WASH sector some of which were successful but very few actually scaled up. So, it occurred to us that a key stage was missing in the development cycle between pilot project and widespread adoption and we called it testing at scale." That's what Millions Saved is all about and I think that's the way to go in the future. |
| [00:18:00] | Finally, I'll just give you one example of a program that doesn't pan out in the end. This is a cash transfer program in Honduras and it increased healthcare utilization but the quality of that care was so poor that actually people's health got a little worse when using. That's the kind of thing you want to know about. Nothing is a silver bullet and it really depends on how it's implemented, trite but true. |
| [00:18:30]<br><br>[00:19:00] | Four lessons that I would take home from this exercise. The first was that the most successful programs were both evidence-based and evidence generating. They started with the best available scientific evidence and once the programs were in implementation they were rigorously evaluated to determine whether the services were delivered efficiently and effectively, and that they were having the intended impact on the final outcome that mattered, health. It's fine to track what we buy or how many doses of vaccine or medicines are sent to a country and we do that a lot in USAID but if you can't measure or report on the program's impact on health outcomes themselves it's hard to say whether the rest of what we do makes sense and there's no way to adjust course to enhance those results. |

[00:19:30] Second, smartly designed programs are targeted to those who will benefit most and use appropriate incentives. Many of the Millions Saved programs gave more funding to the poorest municipalities and districts as in the Brazil Family Health program. Or they identified and reached the most at risk populations as in the [inaudible 00:19:42] India Anti-AIDS program and that strategy has now been replicated more broadly in the AIDS field. Or cash transfers were targeted to those most in need so that really makes a difference because when you're starting from a low baseline you'll have a long way to go up.

[00:20:00] Second, incentives are a powerful tool. For healthcare providers incentives can help motivate greater effort and productivity. If you're not being paid regularly, if no one's checking what you're doing maybe you're not going to do it so enthusiastically. There's always a distribution of performance and productivity among healthcare workers and there are some people who will only show up and do an amazing job but for the rest of us incentives are good. For beneficiaries,

[00:20:30] incentives can motivate people to adopt healthier behaviors and adhere to treatment. As in a tobacco tax that changes the relative prices of products that are not great for you, or helmet laws that instead of just being on the books are enforced, or cash transfer programs that say, "Use the service. It will make a difference for you. It's the worth the cost of getting to that health clinic."

[00:21:00] Third, efficacy is not effectiveness. Efficacy means the product works, it's safe, it can cure a disease in a clinical setting. Effectiveness means that the product is delivered and used in a real-world setting, in a real-world health system with regular people and it has an impact on health that's attributable. I have a friend who's a policymaker in Mexico and he says, "I take a leukemia drug that in the United States saves 95% of children who have leukemia. That same drug in my

[00:21:30] health system saves only 40% of kids." That's the difference between efficacy and effectiveness right there. The system matters. We can't assume that buying an efficacious product and sending it to a country automatically means that the program is going to save lives. We sometimes are still reporting modeled estimates for impact based on the number of things we're buying. It's important but that is insufficient and we now have to go farther.

[00:22:00] The other issue that came out really clearly from Millions Saved is that cost-effectiveness, the cost of producing that next unit of health is rarely assessed in impact of valuation and in at scale programs. Even when it is assessed evaluators many times compare to a do-nothing scenario. This is a huge missed opportunity because it's not convincing. There's always something else going on that people are spending money on. If we want to inform policy and resource allocation we have to

[00:22:30] use counterfactuals that are real life counterfactuals.

Finally, evidence requires its own advocacy and that's a little bit about what Millions Saved is about as well. Of course, we want to use the evidence, we want to generate it to improve outcomes, and we want to improve the quality of our accountability to taxpayers and to Congress. While we selected the programs and policies to feature based on the most rigorous evidence and a valuation available

[00:23:00] we actually wrote Millions Saved to reach the general public and to describe the how of the programs. Their organization, their budgets, their staffing. Decanting rigorous evidence in ways that many users can deploy to best effect is still a challenge but I'm happy to hear a lot of ways that's being done at this conference today and how we could make it happen more.

[00:23:30] Many global health programs has saved millions of lives and prevented disability and some of the most successful programs have been US government supported. My colleague Bill Savedoff and co-author recently did a study evaluating the quality of evaluations in global health. He found that only 4% of global health evaluations at USAID were impact evaluations and that the performance evaluations themselves had quality problems. That means on their own terms they had problems with their analytical and validity reliability.

[00:24:00] We have to realize that this is an agenda that has started but is not complete. I look forward to working with and watching USAID as they implement the amazing evaluation policy that they have. We also have to realize that while diseases like polio and malaria are on the decline they are clear and present dangers still and emerging diseases pose serious and scary threats. Too often we don't know still if our efforts are working because too few programs are rigorously evaluated.

[00:24:30] If there's one thing I'd like to leave you with today it's that health success is possible anywhere, in the very poorest of environments, in the very weakest governance countries if we use the right strategies. I hope to see an elevated role and responsibility for evaluation and evidence at USAID going forward. The gains that we've made over the previous decade give grounds for optimism that in the
[00:25:00] next decade better health policies and programs will mean many more millions saved. Thank you.