



Presents

***IS IT REALLY WORKING?
HIV/AIDS, GLOBAL INITIATIVES, AND
THE PRESSING CHALLENGE OF EVALUATION***

[Transcript prepared from a tape recording]

**Monday, January 23, 2006
2:00 p.m.**

Carnegie Endowment for International Peace
1779 Massachusetts Avenue, Northwest
Washington, D.C.

The Center for Global Development is an independent think tank that works to reduce global poverty and inequality through rigorous research and active engagement with the policy community. This event was made possible in part by funding from the Bill and Melinda Gates Foundation.

Use and dissemination of this transcript is encouraged, however reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License. The views expressed in this transcript are those of the participants and should not be attributed to the directors or funders of the Center for Global Development.

PARTICIPANTS:

Martha Ainsworth, Coordinator
Health and Education Evaluation
Independent Evaluation Group
World Bank

Ruth Levine, Senior Fellow
Center for Global Development

Michele Orza, Study Director
PEPFAR Evaluation
Institute of Medicine

Cyril Pervilhac
Strategic Information and Research
HIV/AIDS Department
World Health Organization

P R O C E E D I N G S

MS. LEVINE: [Proceedings in progress] Development, and I want to welcome you to this event and thank Carnegie for offering the space to us for it.

I would also like to thank Sarah Dean and Myra Sessions for all the work done to put this together.

This event is part of a series of events and analytic work that the Center for Global Development is doing to look at the major sources of funding for HIV/AIDS in the developing world and the performance of those different mechanisms.

Before starting with a few remarks, I just want to highlight that we have another event coming up, the launch of the 2006 Transparency International report, which will be on February 1, at the University Club, pretty early in the morning, at 8:30 a.m.

As many of you may know, the sort of focus on that report is on the health sector. So I think it might be of interest to many of you here.

The evolution of the international response to the HIV/AIDS pandemic can be characterized in many different ways, and here is one. First, there was the doing very little stage, when there was a relatively small amount of financial resources and international attention targeted toward the problem.

Then there was the doing stage, when an increasing amount of financing, activities, analytic work, and programmatic activities were launched and scaled up, often based on very ambitious targets, and a real urgency because there was a growing recognition of the urgent nature of the problem.

And more recently, I hope what we are entering into now is perhaps what we might call the learning by doing stage, where there is a growing acknowledgment that while the problem of HIV/AIDS continues to be urgent and, in many places, increasingly so, it is really moving from being thought of as an emergency to being thought of as something that is a chronic problem, something that the health system as a whole will have to deal with for many, many years, where you are moving from a situation of epidemic to endemic epidemiologic characteristics.

Donors and national governments, local governments and communities are in for the long haul and really have to invest in figuring out what works, both what works in terms of how donors relate to countries and what works all the way down to how specific types of interventions affect the rate of new infections, affect the compliance with treatment and other core factors that determine the public health impact of different programs.

So in this learning by doing stage, we have to think carefully about how it is that that learning is going to take place.

The people we have invited to speak today have devoted their lives, at least in recent years, to thinking about that precise question.

Any evaluation of the very complex programs that constitute the universe of HIV/AIDS prevention, care and treatment programs faces tremendous methodologic challenges, and I am sure we will hear about those.

There are multiple components, which vary by time and by country; diverse contexts in terms of the factors that might affect implementation, as well as the underlying epidemiologic patterns; short time periods of implementation for

some of the programs--so all you can look at are quite process-oriented measures; multiple factors that affect any health outcomes that might be measured; and, a difficulty in establishing a counterfactual, what would have happened in the absence of the program, and I am sure many other problems that will be discussed today, along with what I hope will be some creative solutions and approaches to managing those challenges so that we actually do learn from the programs underway.

Let me introduce the speakers who will speak in the order in which I am introducing them, and then we will have time at the end for questions and answers. We will take those by people lining up at the microphones. So you can think ahead of time about what you might want to ask, and we encourage your questions so that we will get responses, instead of lengthy statements.

So our first speaker will be Martha Ainsworth, who is the Lead Economist and Coordinator for Health and Education Evaluation in the Independent Evaluation Group at the World Bank. And lest you think that that is a new group, it is actually just a new name for an existing group, which used to be called Operations Evaluation Department.

She was the task manager for the recently completed evaluation of the Bank's HIV/AIDS assistance, and all the Bank's activities in this field since the beginning of the epidemic. That will be the marker.

Prior to joining the Evaluation Group at the Bank in 2001, she worked in the Development Research Group in the Africa Technical Department, and you can read more about Martha in the speakers' bios.

Following Martha will be Michele Orza, who is a scholar at the U.S. Institute of Medicine. She joined IOM's Board on Global Health in January 2005 as the study director for the evaluation of the President's Emergency Plan for AIDS Relief, PEPFAR.

Before taking on what I am sure is a challenging assignment, she served as Assistant Director of the Health Care Team at the Government Accountability Office. Again, there are more details in the speakers' notes.

Finally, we will hear from Cyril Pervilhac, who is a monitoring and evaluation specialist for the HIV Department in the Strategic Information and Research Unit at the World Health Organization, and he will be speaking about the World Health Organization's 3by5 evaluation.

So we very much appreciate the presence of our panelists today and we will be looking forward to the questions that you might have at the conclusion of the presentations.

Martha?

MS. AINSWORTH: Good afternoon, everyone. I would really like to extend my thanks to Ruth and to Myra and the others who organized this session. As she mentioned, we have just come out with the first evaluation of the World Bank's HIV/AIDS assistance, but there are a lot more in the planning process and I think we can learn a lot about the pitfalls in designing these things from the experience to date.

Now, since the beginning of the AIDS epidemic through the middle of 2004, the World Bank committed about \$2.5 billion to prevent HIV/AIDS, to treat, and to mitigate the impact in 62 developing countries all over the world.

Has it made any difference and what have we learned from this experience?

These fundamental questions were addressed in the recently released evaluation of the World Bank's HIV/AIDS assistance. It is called Committing to Results, and it was conducted by the Independent Evaluation Group of the World Bank, which I belong to.

It is the first comprehensive evaluation of the Bank's support for HIV/AIDS control at the country level.

Now, among the three evaluations featured at today's roundtable, this is the only one that has been completed and for which there are findings.

For those of you who are interested in the findings, we have the executive summaries outside on the table, with a CD with the full report and all the background materials on it.

I would like to also introduce today Denise Vaillancourt, co-author of the evaluation. Denise, you want to wave your hand? She did a whole lot of the background study work for this evaluation.

Now, this afternoon, the topic is not the findings, but designing evaluations, and I would like to highlight briefly some of the challenges we faced in designing the evaluation, how they were addressed, and some of the lessons we have for evaluations of other programs. Specifically, I would like to highlight these four questions.

First of all, you would probably like to know who the Independent Evaluation Group is; what were the objectives of our evaluation; what was our strategy; and, some of the lessons.

Let me just answer that first question right off the bat. The Independent Evaluation Group, or IEG, is comprised of the Independent Evaluation Units of the World Bank, the International Finance Corporation, and MIGA.

How many people know what MIGA stands for? Okay. It is Multilateral Investment Guaranty Agency. So the IEG is the umbrella for these three agencies and the Evaluation Unit for the World Bank used to be called the Operations Evaluation Department, as Ruth mentioned.

Now, IEG is headed by a Director General of Operations Evaluation, and he is selected by and reports directly to the Board of Executive Directors of the World Bank. So our whole unit reports directly to the Board of Executive Directors, and so does the President of the World Bank. So in that sense, we are independent of the World Bank's management.

We are internal to the World Bank. We all work for the World Bank or the IFC or MIGA, but we are independent of the management of the Bank.

So the objectives of the evaluation. Our objectives really were to evaluate the development effectiveness of the Bank's HIV/AIDS assistance, and, by HIV/AIDS assistance, I mean policy dialogue, analytic work in lending at the country level, since that is where most of the resources have gone, and relative to the counterfactual of no Bank assistance. Establishing that counterfactual is no easy task.

I would like to highlight the fact that the development effectiveness has a special meaning to IEG. Development effectiveness, to us, is a triad of relevance, efficiency and effectiveness.

The second objective is really to identify lessons to guide future activities.

There were two major challenges in the evaluation design. First of all, this is a graph of the commitments approved by the Bank's Board for HIV/AIDS since 1988, Fiscal 1988 through Fiscal 2004. As you can see, there was a steady stream of approvals.

By the way, everything that was committed to a whole project is attributed to the year it was approved. So that lumpiness isn't actually reflecting disbursements.

But what you can see is there is a steady stream over the 1990s and it really picked up at the end. Now, the consequence of that is that about a quarter of the assistance has been completed and three-quarters is ongoing.

So that was the first big challenge, to make it relevant to ongoing projects.

The second, you will see that about a third of all--well, about a half of all the active projects are something called the Africa MAP, which is the multi-country AIDS program in Africa.

The second challenge was this was a new program, with a new methodology, a new approach that hadn't been represented in the closed projects. So it was a challenge to be relevant to the issues faced in this program, even while none of the projects had closed yet. They were all still ongoing. Some of them had only been approved six months before.

These are some of the ways that the approach is different than the standard World Bank approach, but there are documents with even more detail on

this. Basically, there was an emphasis on political mobilization and rapid scaling-up of activities, and the template for this is the National AIDS Strategy, and that was what guaranteed the technical rigor.

So this led us to a strategy of assessing the completed assistance, focusing on some key issues that we knew were being debated in the current portfolio; issues of raising political commitment, of setting strategic priorities in the countries, of the role of the multi-sectoral response, and, I would point out, particularly in relation to the Ministry of Health; the role of non-governmental organizations and civil society; and, monitoring and evaluation. Then, for the Africa MAP, because there was a template governing these whole chunk of projects, we looked at the assumptions and the design of that program in relation to the findings from the portfolio that was closed; specifically, were the assumptions valid; did the project address the main constraints and risks, the design of the MAP; and, then, on implementation to date, what has been implemented; are the objectives being met; and, have the risks that were identified in the design begun to materialize or have they been overcome by the project design.

Then we intend to return later to actually evaluate these projects as they are closed.

The methodology was fairly simple. You are probably all familiar with the results chain that runs from inputs to outputs to outcomes and impacts. I have to underscore that the Bank's inputs in this regard were the policy dialogue, the analytic work, and the lending.

The Bank's performance is integrally tied to the performance of governments. The World Bank always provides its support for government-led

programs. The government may also contract or work with NGOs and civil society, but it is always through governments.

So it is very difficult to disentangle the performance of the World Bank and the performance of the countries themselves, but actually that makes it very interesting because you can learn a lot about the performance of national AIDS programs that way.

Now, in this results chain, aside from the government, we have intermediate outputs. Like political commitments would be an intermediate output. Public, various public policies on AIDS, service delivery, all of these things are intermediate outputs.

Then we have outcomes and household decisions. So decisions on sexual behavior, on condom use, on all kinds of things. Then we have impacts at the level of HIV incidents, AIDS mortality and other indicators.

There is a very complex diagram describing this in Annex A of the final report and I spared you having to look at it. It would have been microscopic on the screen, but it is there and I encourage you to look at it.

Then the later bullets highlight--our strategy was really to document everything along the results chain, including the assistance of other donors. We used time lines extensively, looking at local and international events, when we went into individual countries to look at what happened when and what would be plausible in terms of attribution.

We tried to correlate inputs and outputs with outcomes and impacts, but as you will see, we had major problems because the countries themselves weren't collecting the data that we needed.

In terms of establishing a counterfactual, whenever some information was available, some evidence, we used that, but we also did a lot of interviews of people who were there at the time and asked them what would have happened if the World Bank hadn't come in when it did; would the activities have been funded anyway; would different activities have been funded, and so forth and so on, to try to get at that counterfactual, but it is an art.

Now, we had several building blocks. These are kind of your various inputs that we prepared, literature reviews, portfolio reviews, including looking in the internal World Bank systems to look at what was delivered, surveys, and field work.

I would like to highlight that we did project assessments in Brazil, Cambodia, Chad, India, Kenya, Uganda, and Zimbabwe. Now, these project assessments are the normal bread-and-butter of the IEG. We do project assessments in a quarter of all projects in all the sectors that are completed. So this is kind of an ongoing activity that we do anyway.

Then four case studies, all of them in large countries, Brazil, Ethiopia, Indonesia, and Russia, and they were selected to look at--Ethiopia was ongoing in MAP. Indonesia was an example of a country where the Bank tried to lend early in an epidemic and the project was canceled. They couldn't get the government's commitment.

Russia, where we had five years of policy dialogue before any lending occurred, so that was a closer look at the policy dialogue issue. Then Brazil, where Brazil is into its third HIV/AIDS project, it is one of the major borrowers from the Bank, and Denise was very much involved in Brazil and led the Ethiopia study.

Now, I would like to highlight, I think, the gist of what we want to discuss today, which is the lessons. I would like to highlight four of them here, and then a fifth one just occurred to me as we were hear talking ahead of time, which I will mention.

First of all was this lack of local data. We are very dependent on the availability of local data and we made great efforts to dig in all the project files and all the documentation whenever we went to a country for evidence of outputs that we could link to outcomes; but the poor monitoring and evaluation by the countries and their projects limited the availability severely, and this is something we really need to work on.

It isn't sufficient to say, in HIV/AIDS, that because a national program is financing an activity or an intervention that has been proven somewhere else, that it is going to be effective in that country the way it is being implemented now.

The conditions are too diverse. The political and institutional circumstances are very, very different. One of the main recommendations of our report, one of the only findings and recommendations I will mention today, is to act on local evidence and we really need to collect that evidence on what works in this country at this time, and that constrains severely, I think, our ability to evaluate these programs.

Now, on this issue of monitoring and evaluation, I should mention that we found that it wasn't really a lack of finance for evaluation that was the main problem. There were capacities constraints in a number of countries on monitoring

and evaluation, but the real problem was a lack of incentives, a lack of willingness to link program financing to evaluation.

I am not talking about monitoring here. I am talking about evaluation.

The second point is attribution. There has been a lot of discussion in the international evaluation community about how useless it is to try to attribute any specific outcome in the country to the intervention in any specific donor.

Well, I think they have gone too far. One of the things that we really wanted to know about was the value-added of the World Bank as a player, and I think it would be premature to dismiss this attribution issue. We are a team. We are all working together as a team in the international community and different people have different value-added and comparative advantages.

So we very definitely looked at this and to give you an example, when Denise was evaluating the Brazil HIV/AIDS project, the conclusion there was that had the Bank not come along, Brazil was spending a lot of money on AIDS already and our assistance was only a small part of the total budget; but if the World Bank had not intervened, the program would have been much less focused on prevention and much less focused on the groups at highest risk of contracting and spreading the infection, many of whom are very marginalized.

That is what I mean by the value-added of an individual donor. I hope that in our discussion of attribution that we don't lose track or sight of the need to understand which organizations can intervene most effectively in different aspects.

The third point is the transparency of the findings. For two decades, the international community has been fighting the epidemic, but the urgency of the fight has been the excuse for not evaluating what we do.

So a lot of the decisions now in programs are being made based on notional best practices.

Establishing the facts in these evaluations is very, very important. It takes time, but it is very, very important to put some evidence out there on the table and if people don't like it, then they will have to collect other evidence to show it isn't true. We have to put the evidence out there.

On this evaluation, we attempted to make public and to vet all of our intermediate outputs before the final report was written. We didn't succeed with all of them, but with most of them.

Some of them--Michele, maybe you can hold up--for instance, these project evaluations were public. We also have the four country case studies that were reviewed.

So we have made a very valiant attempt to put whatever evidence we could find out there.

The fourth lesson we learned was the value of an external advisory panel. A lot of the IEG evaluations have these panels. We had a four-person panel, and, basically, their role is to help the evaluations stay relevant, to be a sounding board for ideas we had on the evaluation, and they definitely added a lot of credibility to the findings, I think.

They gave us a lot of advice. We weren't obliged to take it. In the back of the final report is actually an annex which is their statement of what they thought of the report at the end.

So I would recommend an external advisory panel. I think some of us who focus a lot on our own organizations can lose perspective, and they really kept bringing back to us the perspective of the big picture outside the World Bank.

Then the fifth thing that I would mention would be the need--and this is--I have read the terms of reference of a lot of the other evaluations that are coming down the pike, and this was a challenge in ours, was to set manageable objectives for the evaluation.

I have seen so many examples of where more and more and more questions are added. It is impossible to have four or five case studies or field work of some sort and to answer 30 questions. It is just not possible.

So we focused really on did they do what they said they would do; did they support the right things; has it had any effect; that kind of bigger picture. So I think we all want to know what have we got to show for all of this.

So those are five things that I would point to for future evaluations.

Thanks a lot.

[Applause.]

MS. ORZA: I would first like to thank the Center for Global Development for organizing this panel. By inclination and of necessity, the IOM Committee is approaching its evaluation of PEPFAR as a learning process. So it is grateful to have this opportunity to learn from others faced with similar evaluation challenges.

We are all faced with challenges to evaluate evolving, large, complex, multifaceted and far-flung programs. So I thought I would focus on two aspects of the IOM study that create challenges that are perhaps somewhat unique.

One is the IOM's Consensus Committee process and the other is the Congressional mandate and audience for this study.

In the course of briefly discussing these two aspects of our study, I think I can touch on each of the excellent questions that you outlined for today's discussion. Then I will briefly highlight our evaluation plan, which the IOM Committee published just this past October. It is available for free on our website, and I believe that you were kind enough to have copies available on the table outside for people.

Finally, I will briefly update you on the progress of the IOM study. We are right in the middle of it. So, unfortunately, there is nothing I can report to you by way of results, only process.

So, first, the IOM Committee for the evaluation of PEPFAR. Our committee is structured as a core committee of 13 people, with an additional nine people who serve on one of the three subcommittees that correspond to the three major domains of PEPFAR: prevention, treatment, and care.

In addition, we have consultants and liaisons to the IOM Board, for a total of 27 people, all unpaid volunteers. Everyone participates in the work and, in the end, the core committee has to come to consensus on all findings, conclusions, and recommendations.

Our evaluation plan is a product of this process and the report on the results will be, also.

Someone asked me once if it was like herding cats and I realized that it is much more like someone opened all the cages at the zoo. But, seriously, the great strength of this kind of a process is the tremendous breadth and depth of

knowledge, expertise, and experience, not to mention passion, that this group brings to the evaluation.

We have everything from clinicians to modelers, from economists to child psychologists. We have nine members who are not from the U.S. Our Chair is a Mexican, and we have several members who are from three of the PEPFAR focus countries that we are studying.

It can indeed sometimes be a cross-disciplinary, cross-cultural cacophony, but I think the end result is stronger for having had to address such a wide variety of perspectives.

The second set of challenges arises from the origin of our study in a Congressional mandate. The same legislation that authorized PEPFAR, the United States Leadership Against HIV/AIDS, Tuberculosis, and Malaria Act of 2003, also mandates that the IOM evaluate PEPFAR.

Here is the mandate for the IOM study in its entirety, not exactly a detailed scope of work. Responding to a mandate such as this is challenging, for a number of reasons. First, as with all mandates, the client is effectively the entire Congress and even if you focus your negotiations on committees of jurisdiction and other interested parties, in this case, there are a lot of interested parties.

The people who were involved in crafting the legislation and the mandate, the people you are negotiating with now for the terms of work, and the people to whom you deliver the final report may have changed several times.

So the challenges to negotiate terms of work that will be responsive to a wide range of interests, and yet still be doable in your lifetime.

Second, the only two pieces of this mandate that seem clear, that are highlighted, one, that the report should be at the three-year mark and, two, that the legislators were primarily concerned about success, are somewhat at odds with each other.

That is, a report published three years from the passage of the Act cannot reasonably or fairly judge much, if anything, about the success of PEPFAR, and this is partly because of the timetable for the rollout of PEPFAR. Reality always intervenes.

In discussions with Congress, it is clear that they are really interested to know whether PEPFAR is succeeding; are new HIV infections being prevented; are people infected with and affected by HIV/AIDS living longer, healthier lives.

But it is not possible to determine this only a year or so after PEPFAR funding actually hit the ground in the focus countries.

However, Congress also has a need to know about how the program is doing in time to inform its discussions about re-authorizing PEPFAR, which, in this case, would begin around the three year mark.

So what we have negotiated with Congress are three reports intended to meet the somewhat contradictory needs generally described in the mandate.

The letter report that we published in October outlines our plan for an evaluation of PEPFAR implementation. That will be delivered to Congress in the fall of this year, in time to be useful for their re-authorization discussions.

Then, finally, the IOM Committee will provide to Congress another plan for how the ultimate success of PEPFAR could be evaluated after the program is more mature and sufficient data are available.

This is a reproduction of the schematic from our evaluation plan that attempts to show the outline or framework for the entire study all on one page. It is not perfect, but if you only have time to read one page, this is the one you should focus on.

On subsequent slides, it is broken into two halves for easier viewing, but one thing to notice at this point, because it also relates to Congress being the primary audience for this report, is the lack of evaluation jargon.

The basic model that underlies this is the standard one, inputs, process, outputs, outcomes, impact, that many people, including PEPFAR, are using and it can easily be translated into those terms. Those of you who speak evaluationese can probably do it instantaneously.

But Congress generally does not speak evaluationese and doesn't have time for translation. It will be an advantage of our study if it can be understood and appreciated beyond the monitoring and evaluation community, because it will help to make the case for the importance of and usefulness of evaluation to audiences that are able to advocate for and perhaps even fund it, such as the Congress.

If we continue to talk mainly amongst ourselves in the monitoring and evaluation community, I worry that we will not make rapid enough progress toward the day when evaluation is a routine and robust part of any program.

So now I will get off my soapbox and talk a little bit about the evaluation plan.

As we discuss in the letter report, the focus countries are well all of the action is as far as the implementation of prevention, treatment and care activities.

PEPFAR is committed to harmonization and the focus countries are supposed to be leading the way. So they are at the center of the schematic and visits to each of the focus countries are a key part of the IOM Committee's information-gathering activities.

Above the focus countries, the schematic reflects the questions pertaining to how PEPFAR is planned, structured, resourced and managed.

Many of these aspects of the program seem pretty far removed from success, defined in terms of prevention, treatment, and care end points, but they are of intense interest to Congress. For example, the novel structure of PEPFAR and the mandate for coordination at all levels originates in the legislation.

They are also of interest to people running PEPFAR, who have welcomed this external evaluation and hope to learn from it how they can improve their operations. PEPFAR has put a great deal of effort into what it calls the new way of doing business.

More importantly, although far removed from prevention, treatment, and care end points, they are hopefully related to them. I say hopefully, because one of the things that we all struggle with is the paucity of an evidence base demonstrating the connection between some of these inputs and processes and the desired outcomes and impacts.

Below the focus countries, the schematic reflects the questions related to gauging the progress of PEPFAR on its implementation activities and toward its goals of sustainable gains against the HIV/AIDS epidemics in the focus countries.

As with many of the program features from the top half of the schematic, it has been difficult to find metrics for evaluating many of the aspects reflected in this bottom half, and the IOM Committee continues to welcome all ideas for how to evaluate things like harmonization, capacity-building, and integration.

The IOM Committee is also committed to be harmonized with global monitoring and evaluation efforts, and so is relying as much as possible on existing indicators and data sources to avoid reinventing any wheels and duplicating efforts.

At this point, the IOM Committee is deeply absorbed in the focus country visits, represented by the pink box on the flow diagram. Leading up to the publication of the letter report and the launch of the visits to the focus countries, the IOM Committee had four meetings to learn about PEPFAR and to develop the overall evaluation plan and the specific strategy for the focus country visits.

Three of these meetings included sessions that were open to the public and the materials for these sessions are also available on our website.

The focus country visits have been all-consuming since September. Beginning in October, the Committee, in small delegations, has visited Zambia, Rwanda, Vietnam, Tanzania, Botswana, South Africa, and Guyana. Our Mozambique visit just wrapped up on Friday and our Ethiopia team is underway this week. Apparently, things are kind of tense there.

Through the end of February, we plan to visit Kenya, Uganda, Nigeria, Namibia, and, hopefully, Cote d'Ivoire. It looks as though a visit to Haiti, the 15th focus country, will not be possible for security reasons, and so we are planning a virtual visit there. If things don't quite down in Cote d'Ivoire, we may have to employ a similar strategy there.

Even though they have been resource intensive, both the IOM Committee and the PEPFAR missions have found these country visits to be very worthwhile. They represent an extraordinary commitment on the part of the Committee members and the missions, as well, who, as I mentioned, are unpaid volunteers.

Each country visit requires a commitment of at least nine days and many Committee members have done multiple visits.

As soon as the country visits are complete, the Committee will resume other data collection and analysis activities and begin to draft the report to be delivered to Congress in the fall.

Thank you for your attention, and I look forward to the questions and discussion.

[Applause.]

MR. PERVILHAC: Good afternoon. I also wish to thank the Center for Global Development for this meeting and discussion, and I think, as Dr. Levine told us, the learning by doing stage and exercises are quite timely, even in terms of the epidemic and endemic.

At WHO, often, in terms of HIV and AIDS, we speak now of a chronic disease. So it is a welcome discussion in terms of evaluation.

The recent meeting we had of the Monitoring and Evaluation Reference Group, with different partners, in November, stated that the evaluation agenda is really under-represented in relating to implementing programs. So we welcome this discussion.

For this process at WHO of the 3by5 evaluations, it started in August '05. It is going on still until March '06. Also, we don't have any results yet. I just will walk you through the context of the 3by5 and the strategies, as an introduction, and just give you a 3by5 evaluation background, objectives, and scope, so you understand, as well, where we are going.

Then I will walk you through some design aspects and, finally, cover some challenges to evaluation as part of the program lessons learned.

The strategic context for 3by5 stems from the NDG of 2000 and went through UNGAS and led into a WHO HIV/AIDS Plan 2004-2005, which is the WHO strategy worked out here with UN AIDS.

And 3by5 refers to a program of service undertaken by WHO in collaboration with UN AIDS and numerous other partners to support the expansion and access to HIV treatment of three million people in HIV/AIDS in developing countries by the end of 2005, and to simultaneously accelerate HIV prevention efforts, as reflected in that plan.

So the plan is developed and the strategies are developed into five different pillars which the evaluation attempts to address one after the other, and those are under your eyes here. It is, first, related to developing global leadership alliances and advocacy in relation to political commitment and stimulate partnerships towards that treatment, and then responding to urge and sustain

country support with technical assistance at country level and building WHO capacity at country level.

Given that in many offices, we found out that there was very little knowledge about treatment, we had some in prevention and that was quite necessary to build this know-how when we started this in 2004.

Then developing simplified standardized tools for delivering ART was another important strategy that was designed for this 3by5 program, and these were treatment guidelines, training materials, patient tracking and data collection in relation to what we heard about monitoring.

Then strengthening effective, reliable supply and diagnostics with the AIDS Medicine Diagnostic Service pre-qualification program and WHO list of essential medicines, a very delicate and difficult field which involves a lot of policy-making and talk with our pharmaceutical companies and different agencies.

Finally, identifying and reapplying new knowledge and success through operational research, country profiles, case studies, and HIV drug resistance surveillance.

So the background and the objectives of this evaluation were related-- were due because in the grant agreement for a 100 million Canadian dollars we received from the Canadian--from CIDA, from the Canadians as a contribution, it was specified that M&E activities shall be undertaken periodically and at the end of the grant period, and that was after two years, to measure progress and implementation program of work and its impact on scaling-up access to ARV treatment and accelerating HIV prevention at global and country levels.

So WHO, out of this, is also hoping to improve its strategic plans and activities to facilitate the scale-up in countries. So the objectives of the evaluation are here under your eyes, and attempt to answer to the different pillars, the five pillars that I just mentioned and try to review and assess and document these different strategies.

The scope of the evaluation, there are two different scopes. One is related to the core functions of WHO and these, again, are the five pillars I mentioned, and they attempt to answer key questions, such as what has WHO achieved so far, what has it done with the resources at hand, how has it used the resources at central, regional, and country levels, what was its pragmatic technical and organizational contribution.

So it is very much WHO-focused in relation to that donation. The way to work this out is an evaluation at global, that is Headquarters Secretariate in Geneva, regional, and country levels, with a special focus on eight countries.

In addition to that, the scope of the evaluation is to have special focus evaluations, four of them, and this is related--the reason for this is because of highly technical and specialized areas that the team needed to look at, and important questions and interventions for WHO to be proven strategies, and, also, because the evaluation team felt there would be limited opportunities to evaluate these themes during their country visits and needed additional in-depth reviews with the support of other consultants and experts.

So, for example, this IMAI strategy here, why the IMAI strategy? This is Integrated Management of Adolescent and Adult Illness. It was developed and implemented as a package of technical support tools, guidelines, management

procedures, and training activities for scaling up the public health approach to HIV prevention, care, and treatment.

It was scheduled as the main WHO strategy for providing training and technical support to implement the public health approach. This strategy was scaled up in the year 2005 and started in year 2004, and it builds up from the IMCI, what was learned from the child and adolescent health strategy for infectious diseases.

So to move into the few slides I have on the design of this evaluation, they were outlined in the work plan of the evaluation team and discussed at length during the first two steering committees of the team.

The design is largely based on good evaluation practices in terms of reference, and I think you have been handed these terms of reference or they are in the back of the room, if you wish to look at those.

That is, we have looked at the CIDA evaluation guide, evaluation guidelines produced by OECD, the American Evaluation Association guiding principles for evaluators, and we have used, as a practical entry point for us at WHO, the lessons learned from the global polio eradication initiative, who had to evaluate large programs over the past few years.

The design we chose was affirmative evaluation, answering some key questions, such as, for example, what added value has 3by5 standardized and simplified tools for delivering ART contributed to; increasing the number of people receiving ART, health systems strengthening, intensive prevention through scale-up treatment. So these are some of the key questions.

It is not an impact evaluation and it is not an evaluation of the country scale-up program, but of WHO's contribution to reach these targets, and it is participatory, by nature.

A recent phone call I had with the team leader was saying that he appreciated very much this participatory approach, that is, involving the countries and WHO at different levels in this process.

He also mentioned it could have been more strategic if the regional level had been more involved and, also, our strategic information research unit of WHO.

The selection of countries are a little bit common sense in the way that we selected a geographical diversity, with a focus on African countries, different types of epidemics, country presence of key partners, and health care systems and delivery approaches that were different.

At the end, we selected four countries in the Africa region, Burkina Faso, Ethiopia, Malawi, and Kenya, and then Ukraine, India, and Guyana in other countries. The 3by5 has 49 focus countries and that is quite a lot in comparison to PEPFAR; 34 are high burden countries, 15 are regionally strategic countries.

The data collection selected are the standard documentation and data that exist among different partners and WHO, the review of existing studies and evaluation, and field visits. This is a very important step, an enriching step in this exercise, with interviews and surveys of relevant partners in four regional offices.

Seven countries, for political reasons, we could not go to. The team could not visit Ethiopia last November. The team spends two weeks per country. Also, surveying all other countries through questions and e-mails contact that are

independent by going directly to the surveyors, the evaluators. Then special studies and suggested focus evaluation teams, as I mentioned to you.

The standard evaluation practices we use are the same as what we heard from the colleagues in the previous talks; that is, literature review, desk review, interviews, questionnaires, field investigation, special studies.

The government is quite important in the sense that the governance attempts to feed technical matters to this team of seven members that were selected in a competitive bidding process and represent a vast area of technical expertise, and they have met or are meeting four times over this process.

The steering committee meetings, representing key audiences, have taken place twice and one is scheduled next month, and that steering committee is a group, a mixed group of 14 members representing various constituents, bilaterals, ministries of health, technical, people living with HIV/AIDS, activists, evaluation specialists, and UN representatives.

They are different representatives from the 3by5 team members, the WHO Secretariate, and regions and countries in the steering committee. This steering committee provides guidance to the team at different stages of the process.

We assure that a strategic information research team and unit, where I work, is the secretariate for this evaluation.

The budget is approximately 1.2 Canadian dollars, and that is 1.5 percent of the amount that we have for the project, and that covers these standard inputs that we have here in terms of consultation fees, visits, special studies, dissemination processes.

The calendar of activities is a very tight one. It started last August. We had a steering committee in September, a second one in November. The data collection took place in the last trimester of last year. We expect a first draft of the evaluation report at the end of this month, and the final evaluation report in March '06.

The team is presently--just had a meeting, the evaluation team had a meeting and is writing quietly their findings somewhere in Brazil, hiding from all of us, to have peaceful, independent findings.

The overcoming challenges that we found were time constraints, with a very tight schedule. We had hoped for more than eight months, but we had a very slow start, with a two month delay at the beginning, and we also have deadlines because of the reporting to the Canadian Parliament, and this deadline of March cannot be moved further.

We have also found out that the development of the methods and tools is a very time-consuming process.

The team mix is quite good, but we also think that, at the end, based on the experience, what drives the essential selection of good team members is their experience in large-scale evaluations, their experiences in evaluations of a political nature, and being technically competent, and that is a difficult mix to find, including also being available for several weeks over this time period.

The difficulty that the team has found and a big challenge is that in our bilateral projects, we usually function with logical frameworks. We heard a previous speaker tell us about the risk situation and in the case of WHO, the frameworks we function with are office expected results, and that has been very

difficult for the team to try to get the logic of the way we function and the way this project is built into our WHO structure.

The country's availability and constraint is a challenge, because we have found that many countries are willing, but don't have the time to have even a team for one or two weeks.

There are also some political constraints. Sometimes we have even found out there are other evaluations that are more important to them than the 3by5. So it has been difficult to schedule all this.

Guidance without interference is a challenge and the secretariate, the steering committee, and WHO staff has worked hard to respect that.

We have had, for example, the teams going out accompanying the evaluation team and the arrangements, and it is written in the terms of reference that the conclusions and findings are independent in the sense that the WHO staff may not sit in the final evaluation and decisions that are made by the team.

We have found that WHO participation to the process may have helped to own this process and understand what this evaluation is about, and it has worked out quite well.

So protecting independent conclusions is another challenge that we are working on and we are trying to get them to be fully protected to write their findings.

The evaluation, as part of the program, and measurement, the last two or three slides I have. We have had biannual progress reports, because of our mandate and the mandate to the Canadian Parliament; that is, we had an every six month progress report of the 3by5, that you may have heard of, since 2004.

We have the upcoming one for December '05. That is coming after the end of this month, as well, and it is part of the ongoing monitoring.

Using existing process and outcome indicators has been the methodology that we have used. We also had the similar difficulties as the World Bank that sometimes these indicators are not always so good or existent, but the team is--it is dependent of those.

The measurement and assessing key questions is a way that the team is presently functioning. There is no direct attribution of funds, and that is an important point in the sense that it may be a little bit different than my colleague from the Bank; that is, the evaluation encompasses the 3by5 program of work overall, but does not attribute direct financial support for CIDA or any other funding source to specific activities.

However, where possible, the evaluation allocates program costs against inputs and outcomes.

Finally, this evaluation, which took place the last trimester last year, took place while activities were rolling out on a large scale, and this has made the assessments and conclusions all the more difficult.

So the lessons learned, finally, for future evaluations are the following. It has been the exercise of being an affirmative and participatory process, guarantees towards improving its end benefits, with the use of the final results by organizations and various audiences, or so we think so far, but we will tell you a little bit more in a half a year from now, but we believe it will work out with the way the team has functioned.

The team-building process is critical; that is, the evaluation team-building process is critical to harmonize the wealth of diversity of the people who constitute the team.

So it is not like a magic bullet where you tell people here are the tools, go and find this out. There has got to be, also, a very good cement and understanding between the team, and that is a team-building exercise that may need a little bit of time and that cannot be shortcut.

The comprehensiveness of the methodologies and tools is very important and useful, but we do need sufficient time to develop work plans, and I have seen this very nice work plan of PEPFAR, but, also, the appropriate questionnaires for countries.

We have sent questionnaires to the countries, and that is a time-consuming exercise. You cannot send to 49 countries a questionnaire that is unclear or too long.

The mix of countries is essential. The steering committee sometimes, and that is the feeling of the team leader that I spoke to recently for this presentation, felt that it is excellent support, but it may also distract sometimes the evaluation team's need with the questions or queries they have.

Their first feeling now or, let's say, the second feeling, if I may say, it should be a small team working primarily towards driving the evaluation agenda, with expertise in evaluation methodologies.

So that is the kind of steering committee that they find would be most useful, rather than the broad team that we had earlier.

In terms of evaluation perception, it is too early to say. It is still at work, as I said.

Finally, while we think that it is better for the evaluation as an ongoing knowledge development and using and owning nationally the key findings, that is that the learning process can be better worked out if it is owned at the national level, and there has been feedback of the teams at the national level with WHO, but not with the ministries of health, as it principally aimed to WHO.

Finally, the point about the question that I have here towards global partners' evaluation at country level is a little bit of feeling that the team feels that if we are asked individually to go and evaluate our own programs, they feel, okay, fine, if we do it for WHO, but it is difficult, because it is just WHO only and if you are at a country level, you need to understand all stakeholders and partners and try to evaluate those, as well, and that will help you be more objective about what you are evaluating and the questions you are evaluating. So they think that being that no organization works in isolation.

So we need to consider all partners' inputs and contributions to the process, but, of course, this means a more complex methodology. It is more time-consuming and costly. Maybe that is the way of the future.

The last point is the many difficulties we have about indicators and the monitoring and evaluation aspects, and what we heard from the different speakers, that we also felt, for this evaluation, it could be that evaluative research is a way forward; that is, multi-country studies for evidence-based effective package of interventions.

There are talks now between PEPFAR, WHO and Global Fund that were released recently in this publication that tries to explain this a little bit.

It will be a three-year multi-country study, with an additional three to six million U.S. dollars for high prevalence country, building on plan and previous data collection efforts between PEPFAR, Global Fund and WHO, if that takes place, and that could resolve many of the queries we have about trying to understand better outcome and impact.

Another way that could be also the way of the future is what we are doing now with the World Bank on the treatment accelerated program, where we are drawing a learning agenda with research questions and we are documenting those now and we will see what happens in two or three years from now.

So I thank you very much for your attention.

In conclusion, we think we are moving towards universal access by 2010. Global initiatives evaluation are an important stock-taking exercise towards improving and tailoring better strategies in HIV prevention, care and treatment, despite all the challenges that we face.

Thank you very much.

[Applause.]

MS. LEVINE: Great. Thank you very much. If there are questions, and I imagine there are, if you would come to the microphone and, first, state your name and then who you are directing the question to.

Who is going to be the first to pose a question for our panel?

MS. LEWIS: I am Maureen Lewis from the Center for Global Development. It is very interesting to see the different perspectives on how to do evaluations [off microphone].

I have a couple of questions, but my first question has to do with the sensitivity issue [off microphone] to have an internal role in evaluating.

What happened to the [off microphone]? What does that mean? How do you deal with it?

It is a very sensitive subject. A lot of money has been spent in a short period of time and I think a couple of you have mentioned the difficulty of that, but it is something that I think is difficult when you are coming from the inside, but I also commend all of you for doing it.

The second question I wanted to ask was, you are all looking at Ethiopia. Wouldn't it make sense to have, after you finish this, perhaps some integrated look at what has been achieved and what went well, what went wrong, and how do you collaborate and coordinate, in future, either in a better or in a different way, based on the lessons that you have learned?

Because I think it is interesting how you are looking at some of the same questions and some of the same kinds of programs have been promulgated in the same countries.

So it seems to me that there could be some real synergies there from that.

Thank you.

MS. LEVINE: We will take one more question and then I will have the panel respond.

MR. BYRD: My name is Bob Byrd. I used to be in the evaluation business.

This may be a little unfair to our chair, but the Center is launching an initiative on evaluation, and so maybe our chair would be kind enough to kind of indicate, in general, where do you think another approach can augment and what do you propose doing?

Thanks.

MS. LEVINE: That was not a planted question. Let's hear from our panel. I think everybody is curious about this question of political sensitivity and Ethiopia, which I also noted. Cyril?

MR. PERVILHAC: Well, I think the negative findings is a very delicate issue and I think we need not to shy out from this, because if we go out to countries to evaluate those and we have, let's say, the moral responsibility to show, and that is an objective evaluation is positive and negative, we have to do this, and we do this when we go to countries.

Now, vis-a-vis our own institutions, I feel that sometimes we do shy away and we don't exactly react the same way as when we are in countries.

So I think that is certainly something difficult and it needs to be dealt with. Maybe there are ways that we need to explore better in terms of having more process evaluation and not have findings of evaluations at the end that turn off people at different levels.

That is what happens sometimes and I have been trying at WHO to use every opportunity to try to get people to participate to this. In fact, the slides, for example, that we had some preliminary findings from Malawi or others were

presented and criticized during the steering committee meeting, and that has helped a little bit the evaluation team to get the temperature and feeling of what his happening and knowing how WHO will react to these at the end.

But I must confess that looking at some different files at WHO, that indeed, at the end, you have the recommendations that are accepted or not by some panels in-house and that we cannot--that is how it functions.

So at the end, of course, the larger politics may take over.

I'm sorry for my long answer.

Ethiopia, I think, is a very good case in point and I think we need to be collaborating better. I think it is a very good idea to try to see in one country how our findings match and we can use better our findings to work together.

MS. LEVINE: Michele?

MS. ORZA: I think the sensitivities that you were talking about occur on multiple levels, and the first level is at the country level. One thing that is potentially different about the IOM evaluation is that we aren't doing case studies, per se.

We are visiting all of the focus countries in order to get their view on PEPFAR, not to understand how well things are working in Ethiopia or how well things are going in Rwanda.

So there isn't going to be--as a matter of fact, one of the things we have had to do in order to have the countries feel free to talk to us about PEPFAR and how well they are working with the people in the country is assure them that we won't be identifying. So it won't be country X complains that PEPFAR was doing this wrong.

So that level of sensitivity I think we are going to avoid, for better or for worse.

In terms of having aspects of the program that we identify that could be improved, I mean, Congress wants to know about that, because they will be in the process of hopefully re-authorizing this program for another five years and they need to know if it should be authorized differently the second time around.

So I think the way to deal with those is to have constructive recommendations for how to improve the aspects that we have identified that could be improved, but you can't avoid that when the client is Congress. That is what they want to know.

MS. LEVINE: Martha, do you want to reflect, from a slightly different vantage point, on that one?

MS. AINSWORTH: Yes. I don't know how many of you saw this article on scaling up monitoring and evaluation in The Lancet, but they point out this issue that--this recent article in The Lancet pointed out a very important issue, and that is that there is a lot of concern among people who have spent huge sums of money on this, that any kind of criticism will result in a reduction of funding of this effort.

In my view, evaluation is about learning how to be more effective. If you have a valid, relevant objective, you don't stop funding it, because you learn about what makes it more effective and that is our responsibility, is to put out the evidence on what has been effective and what isn't and be very constructive about pointing to the things that we have learned do work and to encourage the data being collected in each and every country.

That said, nobody really likes to be evaluated and, at a certain level, we will never get over the fact that people don't like to be evaluated.

But I think there is traditionally a concern about this accountability. Evaluation is also about accountability, but one goes too far down the accountability route and not enough in the learning route, and that is where I think we run into problems.

When it is only about accountability, then it becomes less--yes.

MS. LEVINE: Let me take advantage of the question that was asked to say that the Center for Global Development is trying to do some work to address the question of why there is systematically under-investment in impact evaluation, which is separate from much of what you all have been involved in, but it was clear from the presentations that it is a missing link in the chain, in causality in many programs, and that you really need to understand what, for example, the public health impact is of all of the funds that are being spent.

You have to really understand, well, what is the impact of the specific types of interventions that are being implemented and scaled up, and in many cases, in a surprising number of cases, that is simply not being--that sort of knowledge is not being generated.

So we are trying to come up with some ideas about what the international community could do to coordinate and fund those types of evaluations, and there is lots of information on our website.

We have got, I think, three people ready to ask questions. Please.

MS. PILLSBURY: I'm Barbara Pillsbury from Social and Scientific Systems.

Any of us who have been involved in these kinds of evaluations know the huge burden that is imposed on people in the countries that we are going to, and many of us are aware that some donors have made commitments to reduce the amount of impact on host country counterparts by visiting teams that come and want to meet with everybody at the management level and make field studies, as well.

I wonder, in the three evaluation processes that you have been reporting on here, if you have used any techniques to specifically minimize the burdens being imposed on host country people, while still being able to get the kind of data that are important for reaching conclusions.

MS. LEVINE: Let's take a couple more questions and then we will get responses.

MS. APTER: Thank you. I'm Felice Apter. I want to thank the panel for being here.

I have a question that probably follows on Maureen's question, in that we are looking at a great set of people who are working with some of the top groups, which have urgency and large amounts of investment, the World Bank, the United States Government, WHO, the Global Fund, which isn't here, and all of these groups take very diverse approaches to their types of investments and that sometimes can lead to a bit of creative tension about who gets to take credit when you get to the point of outcome and impact evaluations.

I am wondering to what extent you are able to address that within your individual evaluations and if there is any organizational, intergroup organization to try to get at some of those more complex questions.

MS. LEVINE: Let's get responses and then we will take two more questions. Martha, do you want to start?

MS. AINSWORTH: Let me focus on the last question, this issue of who takes credit, and I alluded to it in my talk, this issue of attribution.

A prior question is can we show that public policy is having an impact on the AIDS epidemic, period.

Is there evidence to show that what the governments and the donors are doing, even collectively, is having effect, before you get into asking.

What we found was that in a lot of countries, with respect to the World Bank, a lot of the countries where the World Bank has worked, there have been some evolutions, positive trends in knowledge and so forth, and some in behavior.

Maybe it could have been more than that, but we have very little evidence in many of the countries and the links between what was actually done and what those outcomes are. We can't distinguish them from whether this is the normal evolution of the epidemic or a result of public policy.

So the issue of who takes credit, in my mind, doesn't come up in that atmosphere. We have to document better what we are doing and what the outcomes are.

Along that line, in terms of claiming credit, I am constantly frustrated--Martha is griping--I am constantly frustrated by the claims of declining HIV prevalence.

We are in a situation in Africa where HIV prevalence is not the measure, the yardstick for prevention. We have got to look at incidence and we have got to look at AIDS mortality, and this is a problem in all the agencies that I

have encountered and people in the Bank, people outside the Bank, people in the countries, and not just in Africa, of taking HIV prevalence as some kind of yardstick and then claiming that because of our program, it went down.

Well, that is a very sad state of affairs, because prevalence goes down when mortality exceeds the number of new infections, and both could be very high.

So I defer to the others. That was my soapbox.

MS. LEVINE: Any comments on the question of reducing the burden on countries, as well?

MS. AINSWORTH: The burden placed on the people in the focus countries was foremost in the committee's mind when they were coming up with the strategy for collecting data from the focus countries.

At one point, we had entertained the idea of doing some sort of questionnaire or survey in lieu of or in addition to actually visiting, and it was thought that these poor people were probably being sent so many different things that they had to fill out, that an additional one would not be welcome. So we opted not to do that.

We have been trying, with the actual visits, to make them very efficient, to try to take as little of people's time as possible, to try to make the meetings very efficient.

We found that--I mean, Ray is here from the South African Embassy, so he knows how much effort they have put into--the missions have been putting a tremendous amount of effort into hosting us.

But we have really found that people have been eager, because the question we are asking, the bottom line question that we are asking everyone is how has PEPFAR affected your program.

Everybody wants to tell us that. So we have very welcome and the people on the other side have felt like it was a good use of their time.

We are still trying to be conscious of the burden, but it hasn't been a huge issue.

MR. PERVILHAC: On the WHO side, there were the choice of countries that took time and we rejected some countries because they told us no, and the ministry or others said, no, we don't want to, we don't have time or what have you.

So that was one way to try to force the evaluation to some countries, although it was not aimed to the ministry, per se, but more WHO.

Then in the countries, I think the choices of the sites where people would spend a day, two or three days were also carefully discussed with the partners at country level, trying to minimize maybe the burden and, also, the time efficiency.

And, also, the calendar, sometimes we could not go to countries and they said, no, that is not a good timing. So the calendars were moved around, shuffled a hundred times.

Another way, also, were the questionnaires that were sent to the countries, where we tried to shorten them to the minimal questions, not to burden people too much. So these were a few techniques, but I must say that despite this, there is still work to be done for this.

MS. LEVINE: Let's take a couple more questions.

MS. PEARSON: I am Carol Pearson. I am with the Voice of America. I would like the representative from the WHO to please pronounce his name.

MR. PERVILHAC: My name is Pervilhac, Pervilhac.

MS. PEARSON: And your first name is?

MR. PERVILHAC: Cyril.

MS. PEARSON: My question for you is that the overall objective of the independent evaluation is to review the accomplishments and lessons learned during the implementation of the 3by5 in developing countries.

We have got a report that is--the draft is going to be ready by this group holed up in Brazil by the end of this month.

Are there any leaks? If there aren't, or even if there are, you said that we are moving towards global access by 2010.

Could you please talk about that a little bit more?

MS. LEVINE: Why don't we take that, since it is a one-person question.

MR. PERVILHAC: Thank you. Well, I hope I won't be on the Voice of America this evening with my answer, but the audience may be disappointed because I am afraid the leaks cannot be disclosed at this point. If you may be a little bit patient, but you will hear very soon of the progress.

The universal access, I think it may have been for the 3by5, that originally it was a little bit maybe a political WHO move by Dr. Lee during his appointment, but UNAIDS stepped and then other parts.

But I understand the universal access is an attempt to bring all partners on board. I just hope that it will be more successful when I worked 20 years ago in primary health care, that we will be more forceful to reach this target in the years to come.

MS. LEVINE: Maybe evaluations for all by the year 2010.

MS. GRABMAN: I am Genevieve Grabman from the Center for Health and Gender Equity. One of the things my group does is monitors U.S. funded programs in international health.

So I guess my question would be most directed at Michele Orza, but perhaps the rest of you could give some thought to it, as well.

That would be on the tension between the principle of donor coordination, and, one would presume, evaluation coordination, and the thing that I think Ruth touched on in her introductory comments of comparative advantage, and Martha, I think, mentioned value-added, particular donors.

How is it that you would make sure that the results of an evaluation or the findings that one would have would move donors towards a principle of coordination instead of a preemptive idea of comparative advantage?

I think of what PEPFAR is doing right now in terms of there is a perception amongst some policy-makers on Capitol Hill that PEPFAR will have a particular advantage of promoting a particular outcome or a particular methodology in terms of prevention.

Therefore, funding and attention should be focused on this particular methodology for a particular outcome, instead of coordinating and doing a comprehensive approach to, say, prevention.

So I think this gets at a presumption of the comparative advantage before one has actually been found through evaluation and how do you make sure that evaluation contributes not to a presumption of a comparative advantage, but instead to a more comprehensive coordination amongst all donors and, one would argue, amongst all programs and methodologies.

MS. LEVINE: That is perhaps a question only understood by the people in this room, but I think we understand it. Michele, can I ask you to tackle that one?

MS. ORZA: I'm not sure that the IOM can really answer that, from the vantage point from which it is doing the evaluation.

The purpose of the evaluation isn't to ensure anything. It is to describe for the Congress what is going on with this program.

But I think that there is, if you look at the letter report and if you look at the questions that the committee has outlined, there are questions in there that try to get at the tension between how does a program like PEPFAR manage the tensions between saying that they want to be harmonized with the priorities of the country and then also having an agenda for their program and what they want to accomplish.

So what we have at this point are a lot of questions related to those concerns that you have identified, but I think no answers for how to go about ensuring that. I hope that is not too big a copout.

MR. ?: Actually, wait, my question follows up on that, Ruth. So are you going to be able to tell Congress, an individual Congressman, when they ask

you and say, "Hey, how about those abstinence-based programs until marriage," or, "Isn't it a good idea we didn't sort of fund organizations that support sex work."

Congressmen ask these things. Are you going to be able to say things like we don't do harm reduction with PEPFAR, but it would be a good idea if we harmonize? I mean, are you going to be able to tell individual Congressmen, because they ask these questions?

When you go up to Congress, it is like, yeah, great, you did an evaluation, but I want to know what is happening in my favorite country.

Are you going to be able to deal with those tough political questions? Because the authorization language, if you look at it, a lot of the language in that has sort of--some people in public health have issues with a lot of the language there.

So are you going to be able to come out and say, you know, when you do a re-authorization, maybe this wasn't the best use of taxpayer money?

MS. ORZA: I think in the short-term evaluation, that the committee will be able to talk about what the impact on implementation is of a given policy or a given approach or a given set of methodologies that PEPFAR is emphasizing.

They will be able to talk about what kind of progress is PEPFAR making in prevention, treatment and care, and how do those policies or those methods either encourage or hamper that progress.

In terms of the ultimate impact, being able to say that an emphasis on a particular kind of a program or methodology resulted--was more likely to result or less likely to result in the desired outcome, that is going to have to wait for the longer term evaluation.

But I think that there would be a lot of things, and if you look at the questions that are outlined in the letter report, there will be a lot of things that the committee will be able to say about the impact that the way PEPFAR is going about it is having on the implementation of activities on the ground.

MS. NELSON: Joan Nelson, the Woodrow Wilson Center and American University SIS.

We were just essentially talking about donor politics. I am interested in the effects of host country political factors, commitment and the kind of pressures and tensions among various groups.

These strongly influence what is actually done in particular countries. It is true in all fields of activity, not just HIV.

The World Bank evaluation tried to take some of those factors into account somewhat systematically. The group might be interested in hearing a little bit about that.

I happen to know a little bit about that, but I am interested to know whether the other two evaluation approaches that we are hearing about today also tried to take those into account and if so, how.

It is clearly very sensitive and constraints of time bear on that, as on all aspects, and yet it is really important. I would argue that it is probably also important for the planned CDG assessment of what the effects are of the funds and pressures regarding AIDS on public health systems.

So sort of thinking a little bit about how one incorporates those kinds of considerations into evaluation seems to me a useful thing to do, even though it is delicate.

MS. LEVINE: Why don't we, in the interest of time, take the two remaining questions and then give the panel a chance to respond to these three. The last one was one that will take some thought, I think, to respond to.

MS. CURTIS: I am Shawn Curtis, from the Measure Evaluation Project. I have a kind of different question.

Several of the panelists mentioned the difficulty of getting data at the country level, the program data, and how that hinders being able to do a bigger picture evaluation.

I wondered, on their comments, whether they feel that there is progress being made in that area or what more could be done to strengthen the data collection at the country level to save having to go out and collect your own data every single time.

Related to that, as you talk about impact evaluation and we look at things like incidence and AIDS mortality, which are things that are hard to measure, what do you think about what is being done to try and establish the systems now that would then lead to being able to measure those impacts in the long term?

MS. ADATO: My question actually, I think, follows on from that somewhat. Michelle Adato, from the International Food Policy Research Institute.

I am wondering if you could say a little bit more specifically about your analytic frameworks for data analysis.

You talked at the beginning, I think particularly Martha did, about causality and counterfactuals, and I realize that you are using both qualitative and quantitative data, using mostly existing data sets.

I wasn't clear actually the extent to which you were actually designing data collection and using your own data that you could control better.

But I am wondering if you could just say more about your analytical framework and your methods of data analysis, both for the quantitative and the qualitative data, and whether you are able to use or think about using comparison groups.

I mean, obviously not establishing control groups, but to the extent that there might be program rollout with groups that didn't have the intervention, whether that is something that you might end up looking at, and particularly what you learned in the process about what you would do differently with respect to the way in which you designed the evaluation, collected the data, and analyzed it, in particular, in order to be able to better measure specifically impact.

I realize when you are talking about the intermediate variables, it is a different story, but with respect to the other levels, you mentioned household behavior, disease incidence and so on, how you might design things differently, those parts of the process, in order to better answer those questions.

MS. LEVINE: Martha, can I ask you to start?

MS. AINSWORTH: I have to say that I didn't ask Joan to ask that question about political aspects, but she did advise our evaluation and gave us very valuable advice.

Political commitment was one of the major areas of interest and investigation in the World Bank's evaluation. It is an intermediate outcome that was often an objective.

I think I mentioned that IEG always evaluates these activities in relation to whatever objectives they had, and there are usually multiple objectives, but raising political commitment was often an objective of the World Bank's assistance.

So we did find a lot of useful lessons on what can affect political commitment.

We also found that the political commitment in the countries very much mediated what in the package of things the Bank supported were actually implemented.

To give you just one example of that, in a program where there are activities for the general population and there are some activities for people at very high risk of contracting and spreading HIV, like sex workers or injecting drug users or other people with many partners, the general population stuff tends to get done; in a lot of countries, the stuff for the very specific populations does not.

Another thing that tends not to get done is monitoring and evaluation, and that leads into this question from the second person on what we need to do to change this picture.

I mentioned in my talk that we found that it wasn't really that money is a constraint, because we had a lot of examples of where these things were funded, and yet they weren't implemented.

Sometimes it has to do with evaluation capacity, but we think, first and foremost, it has to do with the incentives built in.

Now, these two evaluations are responding to a requirement by a funder to show something at a certain point in time, and that is often not

incorporated into kind of country programs in a very regular way, such that there is kind of ongoing learning by doing.

Until we are willing to say we will fund you for two years and then we want some evaluation and then you reflect on it, redesign the program, and then we continue, unless we make that a condition, I don't think that we will see implementation.

We can train a lot of people, but that doesn't necessarily mean that the evaluation will occur or that the results will be used.

On the issue of causality and counterfactuals, we are really at opposite ends of the spectrum, Ruth and I here. Ruth is talking a lot about randomized control trials, I think, of different interventions.

I live in the world, and I think you all do, too, of places where there is no counterfactual, no simple counterfactual. The whole government program is being evaluated.

So you have to be very crafty in trying to figure out how to figure out what would have happened in the absence of the assistance.

A lot of the projects that the Bank supported had pilot interventions, but they were often scaled-up without being evaluated. If those things were actually evaluated before they were scaled up, then one would have some kind of comparison group, and that gets back to the incentives.

They are often piloted for political and technological feasibility, but they are never actually evaluated with respect to their costs and effects.

One of the disappointments was that the four case study countries we were looking at, the impact of all the Bank's assistance, not just its lending, but the

analytic work and the policy dialogue, and we chose big countries for the reason that we thought that--take the states of Brazil--that we would see some variation across the states and different degrees of implementation of the program and be able to work off of that to come up with some sense of relative project inputs and outputs and then different outcomes, but, unfortunately, there just wasn't data disaggregated enough.

I mean, we were hoping to do the same thing in Ethiopia, and we weren't able to. So that was a shame.

But I would like to emphasize the second questioner's point on we have to, on this issue of looking at incidence and AIDS mortality, we really can't wait to put into place these mechanisms to measure the important indicators that we need, and it is kind of shocking to me that this isn't happening to a greater extent.

We have to put those things into place now. If we really want to understand whether these programs work, we have to know about AIDS mortality and HIV incidence.

MS. ORZA: I think that in any case, but particularly because of the commitment to harmonization, a program like PEPFAR has to be flexible enough to be able to take on the shape of whatever country it is trying to operate in, and the leadership at the country level, the political commitment or lack thereof is one of the factors that the committee understood would have an impact on how well PEPFAR was able to function in a particular country.

I think that the questions that will get at what is going on there are in the--again, all I have at this point are questions. I don't have any answers, but I think they are in the coordination section and they are in the harmonization section.

So what will come out there will be the impact of working in a country with a particular set of leadership challenges on the implementation of a program like PEPFAR.

One of the things that PEPFAR has said it is going to do is to increase the capacity of these countries to deal with their epidemics, and one important capacity that most of them need is monitoring and evaluation capacity.

So there are questions both in the PEPFAR management section and in the capacity-building section that relate to what activities is PEPFAR undertaking to help strengthen the monitoring and evaluation capabilities of these countries, so that eventually we can stop bemoaning the lack of data and have some data to work with.

That is one of the things that the committee will be looking for, the extent to which PEPFAR is effectively supporting that.

Then in terms of analytic frameworks and having a comparison group, for the short-term evaluation, the question is really how is PEPFAR doing, not how is PEPFAR doing relative to other people, but really how is their implementation going with respect to their own milestones and their own goals.

I think where we potentially would consider a design like that and have a different sort of a framework is when we can do the longer term evaluation and really be looking at impact to say how much of that really was attributable to the way that PEPFAR was going about it, how many of those results really can PEPFAR claim credit for, but in the short term, it is not on the radar screen.

MR. PERVILHAC: Thank you. On the question of the political

commitments, I think maybe--I mean, I would like to see a tool that can help us to measure and assess this systematically in countries.

It is yet to be seen, but maybe it is there. But I can just read to you a recommendation that came out of the 3by5 evaluation in a country I would not mention, but it really relates to policies.

It says, "An observation to WHO would seem to advocate within the government for broad developmental approaches to strengthening treatment and care, especially in the fields of nutritional aspects and agricultural policies."

So I think that is the kind of recommendation that a team that works in the country, even without looking specifically at political commitments, can grab and come up with some very concrete and excellent recommendations.

Another domain that I think you are asking on the monitoring and evaluation, is progress being made. I think, yes, progress is being made and it is very encouraging.

I was recently in Mali and, at the national level, they did come up with, I think, like 8,000 people in treatment.

Of course, the problem we have is what are the details of this, what is the number of dropouts, what is the number of people switching drugs.

You see, we have to think, when we--some of you were present out there in the field when, 20 years ago, we had the large program called African Child Survival Initiative, and it took us several years of large efforts to build up our immunization systems.

Here we are talking, for example, of treatment, care, psycho-social, and we want this to happen within two or three years. No. It takes time. It took

us five years of intensive efforts to build these efforts, and they were much simpler than these systems are we are looking at here.

So we just have to be patient and I think we are making progress, and I think projects such as measure evaluation is very useful.

The question on impact is a very difficult one. My colleague, Dr. Mead Over, would be better to answer this, but I think we said at the end that multi-country studies, special studies could try to answer impact with a special design, and these are not part of routine information.

Thank you.

MS. LEVINE: Well, I think that we have heard this afternoon from three people who represent a really key function in public service, generally, which is the function of evaluating, with a clear eye toward what is the best way for public and private funds to be spent to achieve the goals that we all share.

These are individuals who represent a profession that runs counter to policy-based evidence-making; that is, generating evidence and analyses to back a particular advocacy position or a policy, and instead they are supporting evidence-based policy-making, which I think is something that is increasingly recognized as having a real value in the hard decisions and programs that we are working on today.

So I really appreciate the speakers and I know that you don't get much glory in your work. I hope that you have enjoyed the opportunity to present your ideas today, and I very much thank you for coming to this event and I hope that we will see you at future CDG events.

[Applause.]