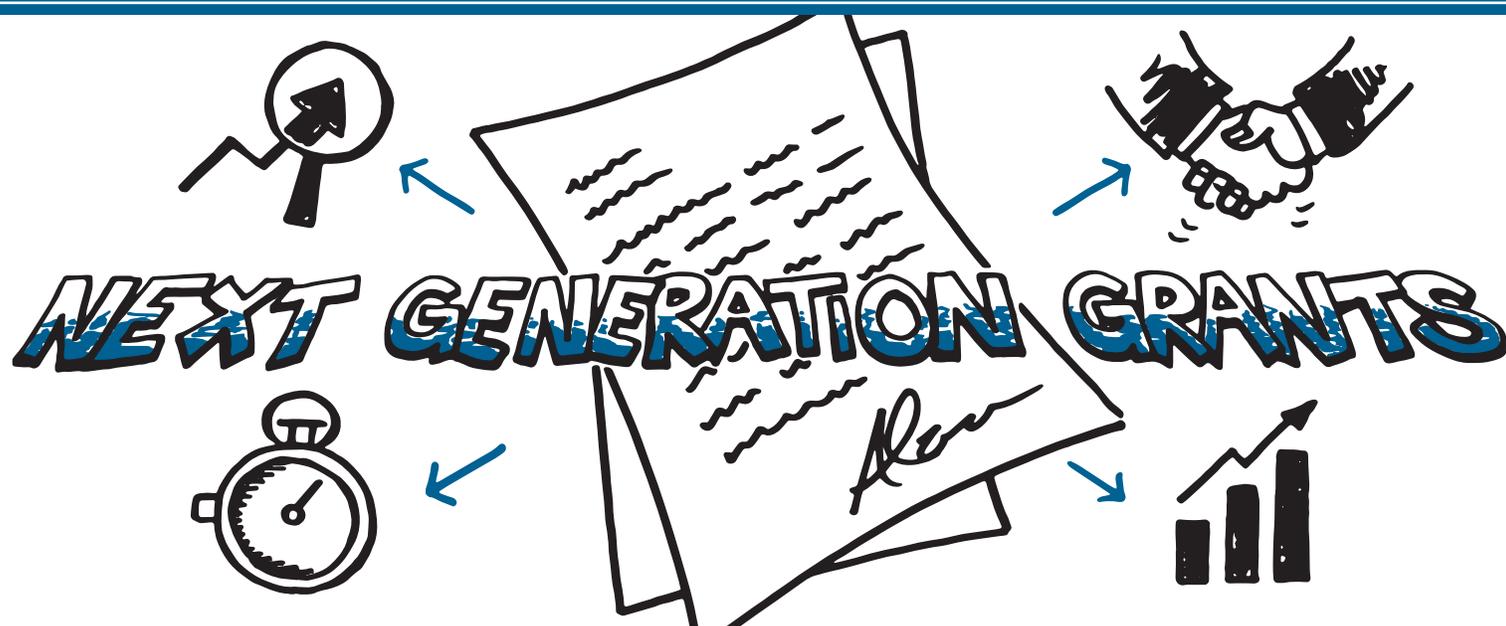


Aligning Incentives, Accelerating Impact

Next Generation Financing Models for Global Health



A Report for the Center for Global Development Working Group
on Next Generation Financing Models in Global Health
Amanda Glassman and Maria Kirova, Co-chairs

Report authored by Rachel Silverman, Mead Over, and Sebastian Bauhoff

CENTER FOR GLOBAL DEVELOPMENT

Aligning Incentives, Accelerating Impact

Next Generation Financing Models for Global Health

A Report for the Center for Global Development Working Group
on Next Generation Financing Models in Global Health
Amanda Glassman and Maria Kirova, Co-chairs

Report authored by Rachel Silverman, Mead Over, and Sebastian Bauhoff

This document is a final report of the Center for Global Development (CGD) Working Group on Next Generation Financing Models in Global Health. The report draws from working group discussions and a series of background papers commissioned by CGD and authored by Dave Burrows and Lou McCallum; Elvin Geng and Nancy Padian; Rachel Silverman; Liam Wren-Lewis; and Han Ye. The authors are additionally grateful to other participants in an April 2015 technical workshop that helped inform this work—Arnab Acharya, Stefano Bertozzi, Jed Friedman, Paul Gertler, Kate Harris, William Jack, Hélène Roger, Naline Sangrujee, and William Savedoff—and to participants in an October 2015 meeting on malaria indicators—Ibrahima Diallo, Mamadou Lamine Diouf, Erin Eckert, Scott Filler, Christie Hershey, Medoune Ndiop, Deepa Pindolia, Arsène Ratsimbao, Richard Reithinger, Estifanos Shargie, and Steven Yoon. They extend additional gratitude to John Osterman and Lauren Post for their help in preparing this report for publication. The authors are grateful to the Bill & Melinda Gates Foundation and UK Department for International Development for financial support of this working group.

© Center for Global Development. 2015. Some Rights Reserved.
Creative Commons Attribution-NonCommercial 3.0

Center for Global Development
2055 L Street NW, Floor 5
Washington DC 20036

www.cgdev.org

ISBN 978-1-933286-94-5

Editing, design, and production by Communications Development
Incorporated, Washington, D.C., and Broadley Design, Bristol, UK

Working Group on Next Generation Financing Models in Global Health

Working Group co-chairs

Amanda Glassman, Center for Global Development

Maria Kirova, the Global Fund to Fight AIDS, Tuberculosis and Malaria

Working Group members

David Barr, The Fremont Center[†]

Sebastian Bauhoff, Center for Global Development*

Alexander Birikorang, The Global Fund to Fight AIDS, Tuberculosis and Malaria*

Olga Bornemisza, The Global Fund to Fight AIDS, Tuberculosis and Malaria[†]

Michael Borowitz, The Global Fund to Fight AIDS, Tuberculosis and Malaria

Robert Brinckman, Formerly of the Clinton Health Access Initiative[†]

Dave Burrows, APMGlobal Health*

Jason Clark, The Global Fund to Fight AIDS, Tuberculosis and Malaria[†]

Lucica Ditiu, Stop TB Partnership[†]

Rajesh Divakaran, The Global Fund to Fight AIDS, Tuberculosis and Malaria[†]

Erin Eckert, President's Malaria Initiative[†]

Peter Godwin, Independent*

Matias Gomez, The Global Fund to Fight AIDS, Tuberculosis and Malaria

George Gotsadze, Technical Review Panel, The Global Fund to Fight AIDS, Tuberculosis and Malaria

Katherine Guerra, Clinton Health Access Initiative*

Kate Harris, Bill and Melinda Gates Foundation[†]

Jacqueline Huh, Stop TB Partnership*

Ashaque Husain, National Tuberculosis Control Program, Bangladesh[†]

Pablo Ibararán, Inter-American Development Bank[†]

Harinder Janjua, The Global Fund to Fight AIDS, Tuberculosis and Malaria

Sebastien Lenelle, Office of the Inspector General, The Global Fund to Fight AIDS, Tuberculosis and Malaria[†]

Julia Martin, Office of the Global AIDS Coordinator[†]

Linden Morrison, The Global Fund to Fight AIDS, Tuberculosis and Malaria*

Cynthia Mwase, The Global Fund to Fight AIDS, Tuberculosis and Malaria[†]

Regina Ombam, National AIDS Control Council, Kenya[†]

Mead Over, Center for Global Development

Nancy Padian, Berkeley School of Public Health

Hoangmai Pham, Center for Medicare and Medicaid Innovation

Ferdinando Regalia, Inter-American Development Bank*

Yuna Sakuma, Center for Global Development

Motoko Seko, The Global Fund to Fight AIDS, Tuberculosis, and Malaria*

George Shakarishvili, The Global Fund to Fight AIDS, Tuberculosis and Malaria*

Alexandra Shields, Clinton Health Access Initiative

Rachel Silverman, Center for Global Development

Pavlo Smyrnov, International HIV/AIDS Alliance

Tracy Staines, The Global Fund to Fight AIDS, Tuberculosis and Malaria*

Kate Thomson, The Global Fund to Fight AIDS, Tuberculosis and Malaria[†]

Veronica Walford, World Health Organization

Lisa Waugh, Clinton Health Access Initiative

David Wilson, World Bank

Aneta Wierzynska, The Global Fund to Fight AIDS, Tuberculosis and Malaria*

Nathalie Zorzi, The Global Fund to Fight AIDS, Tuberculosis and Malaria

[†] Attended first meeting only.

* Attended second meeting only.

Contents

Disclaimer	vii
Preface	viii
Glossary	ix
Executive summary	xi
Chapter 1 Why this report? Why now?	1
Big ambitions, limited leverage and visibility: Why the principal-agent relationship and contract theory matter for global health financing	2
Next generation financing models: A new approach to an old problem	3
Chapter 2 Learning by doing: An evolving approach to incentives at the Global Fund and elsewhere	5
Performance-based financing: Ahead of its time, but diluted incentives and an uncertain future	5
From pilots to strategy: The Global Fund explores new financing models	6
Balancing promise against risk	6
Looking outside: What can be gained by getting incentives right in global health?	7
Chapter 3 Changing the basis of payment: Why, where, and how much?	10
What does it mean to change the basis of payment?	11
Why might the Global Fund and its Principal Recipients want to change the basis of payment?	11
In what contexts would changing the basis of payment be desirable—and where would doing so be feasible, either in full or in part?	12
A modular approach to changing the basis of payment	18
Chapter 4 Getting the incentives right: Next generation financing models	21
Starting simple: Fixed price models	21
Hedging your bets: Alternatives to simple fixed price mechanisms	24
Protecting communities, safeguarding rights	26
Chapter 5 How you know: Verification, assurance, and assessment	28
Why verify?	28
Tradeoffs and credible threats	28
Verification in practice	29
Learning while doing: Assessment and evaluation	31

Chapter 6	Country case studies	35
Case 1:	HIV in Country A	35
Case 2:	Tuberculosis in Country B	37
Case 3:	Malaria in Country C	38
Case 4:	HIV in Country D	40
Chapter 7	Recommendations for the way forward	43
Notes		47
Bibliography		48
Boxes		
3.1	Limited-capacity Principal Recipients and next generation financing models	13
3.2	Contractible indicators for malaria impact: A unique challenge	16
3.3	Increasing Principal Recipient prefinancing: A good idea for the Global Fund	19
4.1	Real world example: The Salud Mesoamerica Initiative	22
4.2	Real world example: Cash on Delivery for education in Ethiopia	22
4.3	A nonlinear payout schedule to elicit information on efficiency	24
4.4	Dynamic fixed price models in the real world	25
4.5	Real world example: Travel per diems	25
4.6	Real world example: Medicare shared savings program	26
5.1	Verification in the Salud Mesoamerica Initiative	30
Figures		
1.1	Value for money domains for global health funders	1
1.2	Next generation grantmaking within the results-based aid universe	4
3.1	The Global Fund grant portfolio at a glance (active grants as of mid-2015)	10
3.2	Grant designs along the input financing to payment for results continuum	19
3.3	Four hypothetical grant models	20
4.1	Linear and kinked payment schedules	23
5.1	Verification scenarios and associated disbursements to the Principal Recipient	32
5.2	Sample assessment framework—Hypotheses and risks throughout the grant cycle	32
Tables		
3.1	Potential bases for payment along the results chain	11
3.2	Hypothesized benefits of changing the basis of payment	12
3.3	Sample contractible indicators	14
3.4	Sample noncontractible indicators	15

Disclaimer

This report is a product of the Center for Global Development. Its content is based on the deliberations of a working group comprising individuals from the Center for Global Development, the Global Fund, governments, technical partners, civil society, other aid agencies, academia, and multilateral development banks. Members of the working group served in their individual capacities, not as official representatives of their organizations.

All members of the working group have had the opportunity to review this report. However, working group members, including staff of the Global Fund, do not necessarily endorse all components of this report, nor do the contents of this report constitute a policy commitment by the Global Fund or any other party. Views expressed herein can be attributed to the authors alone. All errors and omissions are those of the authors.

Preface

Performance-based financing, results-based financing, outcome-based aid, Cash on Delivery, and now Next Generation Financing Models: does a rose by any other name really smell as sweet? The alphabet soup of names and acronyms grows longer with each passing year. Yet the proliferating glossary heralds a growing and common-sense donor consensus: incentives matter. Scarce aid dollars will go farther if they pay for better health, not just for doctors' salaries and syringes.

As one of the world's largest health funders, The Global Fund to Fight AIDS, Tuberculosis and Malaria (the Global Fund) has long aspired to link its funding to results, and has elevated performance-based financing as a core value since its 2002 launch. But performance-based is easier said than done—a challenge noted by my colleagues at the Center for Global Development (CGD) over a decade-plus of research and analysis. As early as 2004, Steve Radelet highlighted challenges in measurement and enforcement of performance-based financing at the Global Fund; then, in 2006, he led a working group that urged the Global Fund to strengthen the performance-based financing mechanism. In 2010, under Nandini Oomman's leadership, CGD's HIV/AIDS Monitor Initiative noted weaknesses in verification of self-reported performance—the ostensible basis for the performance-based financing system. Most recently, CGD's Global Health Policy team led an extensive working group on Value for Money at the Global Fund. Their analysis showed little systematic relationship between measured results and actual disbursements under the performance-based financing system, leading them to recommend that the Global Fund tie at least a portion of its funding directly to measured and verified progress against the most important health indicators.

Building on this enormous portfolio of CGD's prior work, this report suggests a new direction for the Global Fund's financing model—one with better incentives, enhanced accountability, and

a renewed focus on performance verification. Next Generation Financing Models linking results to disbursements, including my own proposal for Cash on Delivery aid, have enjoyed growing enthusiasm among global health and other development funders. But as my colleagues Rita Perakis and William Savedoff report in a recent policy paper, their implementation has thus far been cautious and slow. Funders face both real and perceived barriers to the rollout of innovative financing models, coupled with little precedent on how to address those challenges.

Tailored to the donor audience, this report thus offers practical guidance on translating new financing modalities from proposal to practice. Starting with the high-level principal-agent problem and then zooming in to the nitty-gritty of grant design, the report marries cutting-edge, Nobel-Prize-winning economic theory with the operational realities of a large multilateral funder. It answers funders' frequently asked questions about Next Generation Financing Models, including how to select indicators, how to design a payment mechanism, and how to rigorously verify performance.

I am pleased and encouraged that this report has emerged from a co-chaired working group with the Global Fund itself—a fruitful collaboration and exciting opportunity for translating the policy proposals presented here into concrete and actionable changes at a major global health funder.

The Global Fund is evolving—and I am proud that CGD is a partner in this process. I urge the Global Fund Board to support this new direction, and hope that the Secretariat will closely follow the technical recommendations put forth in this report. I look forward to following these developments and reporting on the Global Fund's progress in the years to come.

Nancy Birdsall
President
Center for Global Development

Glossary

- basis of payment** The theoretical and practical basis for Global Fund disbursements, for example, what the Principal Recipient must do and document to justify those disbursements. In traditional grants, the basis of payment is allowable expenses, incurred and documented.
- Cash on Delivery** A specific instrument under the broader umbrella of fixed price payments, coined by Birdsall and Savedoff (2010) of the Center for Global Development, where donors pay for measureable and independently verified progress on specific outcome measures.
- contractible indicators** Indicators that can serve as the basis of payment for results payments. These indicators must be at least partially influenced or under the control of the Principal Recipient or health system, measureable, independently verifiable, and a direct proxy or on the direct causal chain to a meaningful and important health outcome. They must not incentivize coercion or human rights abuses.
- cost reimbursement** A grant or portion thereof where the Global Fund would reimburse the Principal Recipient for its documented expenditures.
- Development Impact Bond** A financing arrangement wherein a social investor provides up-front funding for a development project—and is subsequently repaid by donors or a country government if the project achieves verifiable progress against pre-agreed outcomes.¹
- ex ante** Forecast or done prior to an event of interest. In this report, ex ante refers to the period before agreement and implementation of the grant.
- ex post** Calculated or done based on the actual results during the period of interest. In this report, ex post refers to decisions made after results are observed during grant implementation.
- fixed price** A way to structure results payments in which the Principal Recipient is paid a fixed unit price for every unit of verified performance or improvement; unit price can be adjusted for quality. This umbrella category includes instruments like Cash on Delivery and output-based aid.
- fixed price/cost reimbursement menu** A way to structure cost reimbursement and results payments in which the Principal Recipient is offered a choice between fixed price and cost reimbursement; the Principal Recipient does not have to decide between the two choices until the end of the payment period, when it knows which option will be more advantageous.
- Fund Portfolio Manager** The Global Fund’s Geneva-based grant managers.
- information asymmetry** A situation in which one party to an agreement has more information than the other, putting the latter party at a disadvantage.
- linear payout schedule** A payment schedule that offers a single, constant price for each unit of output, so that the total payment is just the number of units multiplied by the price per unit.
- Local Fund Agent** In-country accounting or consultancy groups contracted by the Global Fund to oversee Principal Recipients’ financial management, compliance with Global Fund policies, and operational performance.
- New Funding Model** The Global Fund’s current funding mechanism—fully operational since 2014.
- next generation financing models** Agreements that create an explicit and enforceable contract between funder and recipient, wherein both parties agree that the financial relationship will be directly tied to the achievement of mutually important, realistic, and measurable gains in the provision of health services or population health.
- nonlinear (or kinked) payout schedule** A payout schedule in which the per unit payment varies depending on the level of achievement, for example, to accommodate varying economies of scale or to enhance rewards over more difficult output ranges.
- participation constraint** A tenet of contract theory that states that for a grant mechanism or design to work, the agent must agree to take part.

- payout schedule** A pre-agreed rule that specifies how different levels of Principal Recipient achievement will elicit different levels of payment.
- performance-based financing** A component of the Global Fund's traditional grantmaking model that allows for the adjustment of the grant ceiling on the basis of a holistic assessment of grant performance.
- power (of a grant model)** The extent to which the design of a grant can motivate efficiency gains, which is a direct result of the Principal Recipient's ability to benefit from those efficiency gains (for example, to keep residual savings).
- prefinancing** Payments made in advance by the Principal Recipient, with the possibility of later recouping those funds from grant payments, either through cost reimbursement or results payments. Money for prefinancing could come from a variety of sources including the Principal Recipient's reserves or current operating budget, a transfer from the country government, a loan from a private or multilateral bank like the World Bank, or even a private investor.
- principal-agent relationship** A relationship where an organization, known as the principal (for example, the Global Fund), agrees that a second organization, known as the agent (for example, the Principal Recipient), will act on the principal's behalf and pursue the principal's objectives, which may be more or less similar to those of the agent.
- Principal Recipient** An organization that enters into a direct grant relationship with the Global Fund, whereby it receives funding directly from the Global Fund and is responsible for implementing the grant agreement. A Principal Recipient may have one or more subrecipients.
- results-based financing** An umbrella term for a broad range of financing instruments that attempt to connect at least a portion of payment to the verified achievement of results.
- results payments** A grant or portion thereof where the Global Fund would pay the Principal Recipient ex post on the basis of verified outputs, outcomes, or impact.
- shared surplus** A payment model where the funder would project the expected costs of reaching a prespecified health goal, and allow the recipient to keep a portion of the surplus if the recipient achieves the health goal at a lower than expected cost.
- subsidy payments** Fixed price payments that subsidize the achievement of results—for example, fund some portion (but less than 100 percent) of the actual cost.
- substitute payments (for input financing)** Fixed price payments that substitute for input financing—that is, payments that fund the entire cost of service delivery.
- supplemental payments** Fixed price payments provided on top of input financing—either by the Global Fund itself, the government, or another donor—that incentivize the Principal Recipient to efficiently translate existing inputs into better health results.
- Technical Review Panel** An independent panel of technical experts that reviews proposals to the Global Fund for technical merit and makes funding recommendations.
- verification** A process to check the robustness of reported results and ensure that they offer a sound basis of payment.

Executive summary

Founded in 2002, the Global Fund to Fight AIDS, Tuberculosis and Malaria (the Global Fund) is one of the world's largest multilateral health funders, disbursing \$3–\$4 billion a year across 100-plus countries.² Many of these countries rely on Global Fund monies to finance their respective disease responses—and for their citizens, the efficient and effective use of Global Fund monies can be the difference between life and death.

Many researchers and policymakers have hypothesized that models tying grant payments to achieved and verified results—referred to in this report as next generation financing models—offer an opportunity for the Global Fund to push forward its strategic interests and accelerate the impact of its investments. Free from year-to-year disbursement pressure (like government agencies) and rigid allocation policies (like the World Bank's International Development Association), the Global Fund is also uniquely equipped to push forward innovative financing models. But despite interest, the how of new grant designs remains a challenge. Realizing their potential requires technical know-how and careful, strategic decisionmaking that responds to specific country and epidemiological contexts—all with little evidence or experience to guide the way.

This report thus addresses the how of next generation financing models—that is, the concrete steps needed to change the basis of payment from expenses to something else: outputs, outcomes, or impact. For example, when and why is changing the basis of payment a good idea? What are the right indicators and results to purchase from grantees? How much and how should grantees be remunerated for their achievements? How can the Global Fund verify that the basis of payment is sound—that the reported results are accurate and reliable and represent real progress against disease control goals? And what is needed to protect communities against coercion or other human rights abuses, ensuring that these new incentives do not drive unintended consequences?

A move toward next generation financing models includes real risks. Yet those risks must be understood compared with the very real risks of maintaining the status quo: the risk of not achieving maximum impact, the risk of not having full visibility on whether Global Fund monies ultimately reach their intended beneficiaries,

and the risk that ultimate goals will never be attained if the Global Fund and the recipients of its funds do not fully align around shared objectives. The opportunity cost of the conventional approach may affect whether people in low- and middle-income countries are freed of the terrible costs of AIDS, tuberculosis, and malaria.

This report is divided into four parts. Part I offers a conceptual framework that explains why traditional grantmaking often gets the incentives wrong, why that matters, and how next generation financing models might offer a way for the Global Fund and other health funders to increase the value for money of their investments. It also describes the growing use of incentives at the Global Fund and elsewhere, including the current incentives embedded within Global Fund grants.

Part II discusses contexts where a move to next generation grant models could drive faster impact or other benefits for the Global Fund and describes the technical elements and design choices required to bring them to life. It focuses on four core questions:

- Where and why might the Global Fund consider the introduction of new financing models?
- What should the Global Fund pay for—that is, what are the characteristics of a contractible indicator, and which indicators for AIDS, tuberculosis, and malaria fit that bill?
- How much and how should the Global Fund pay—that is, what makes for a good payout schedule, accounting for varying local contexts?
- How can the Global Fund ensure that the basis of payment is sound—that is, how can the Global Fund verify the accuracy of reported performance and deter attempts at fraud or manipulation?

Part III “puts it all together,” providing four illustrative examples of how next generation grants could be designed and applied within the Global Fund's grant portfolio. It provides and applies a multistage framework for their design and implementation. During the preparatory phase, the Global Fund must understand the situation and context; define the policy objectives; and review strategic considerations, risks, and constraints that may influence the choice of financing model. During the design phase

the Global Fund must define key details of the grant agreement, including the portion of funding allocated to results payments, and the contractible indicators, payment mechanism, and verification strategy used for results-based payments and define all outstanding issues to be resolved through negotiations with the Principal Recipient. Following the negotiation, the Global Fund must review the entire grant model and ensure that it responds appropriately to the policy objectives, addresses contextual constraints, and mitigates possible risks. Finally, during and after implementation the Global Fund must assess and evaluate the performance of the model in practice.

Part IV concludes with seven recommendations for the way forward—that is, the operational and strategic steps that the Global Fund will need to take in order to realize these designs and maximize their potential benefits:

1. Secure strong Board and Secretariat commitment through inclusion of next generation grants as a key priority within the next Global Fund Strategy (due to be presented to the Board for approval in April 2016).
2. Leave no room for ambiguity: ensure that next generation grant agreements stick to their agreed disbursement protocols—against progress on independently verified results.
3. Reflect the needs and requirements of next generation grants in relevant related policies, including the allocation formula, counterpart financing requirements, sustainability framework, and differentiation initiative.
4. Reflect the needs and requirements of next generation grants in the guidance and terms of reference given to key Global Fund bodies, including the Technical Review Panel, Country Coordinating Mechanisms, and operational divisions within the Secretariat.
5. Assure Global Fund and Principal Recipient access to needed expertise and resources to design and operationalize next generation grants, with particular attention to performance verification.
6. Revise Key Performance Indicators to accommodate differences in the management and evaluation of next generation grants.
7. Evolve financial management policies to accommodate less predictable cash flow and reduce restrictions on the use of funds.

Part I

Why next generation
financing models?

Chapter 1

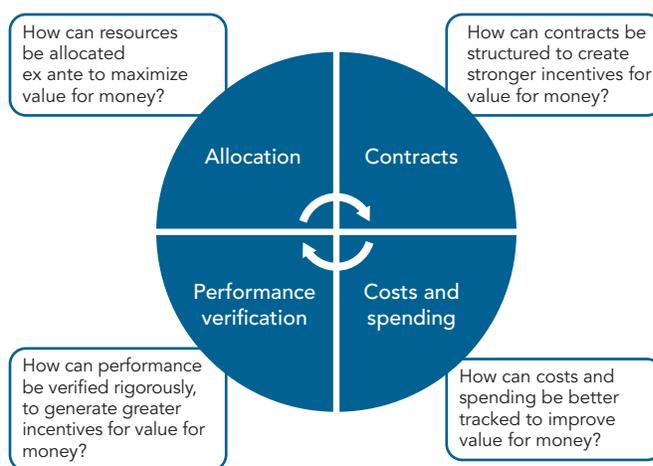
Why this report? Why now?

Founded in 2002, the Global Fund is one of the world's largest multilateral funders for health, describing itself as “a 21st-century partnership organization designed to accelerate the end of AIDS, tuberculosis and malaria as epidemics.”³ The Global Fund disburses \$3–\$4 billion a year across 100-plus countries.⁴ Many countries rely on those funds to finance their respective disease responses—and for their citizens, the efficient and effective use of Global Fund monies is often the difference between life and death.

Recognizing how much the Global Fund matters, both for its direct beneficiaries and for the world as a whole, the Center for Global Development (CGD) released a report in 2013 entitled *More Health for the Money: Putting Incentives to Work for the Global Fund and its Partners*.⁵ The report suggested four domains where the Global Fund could stretch its scarce dollars to maximize its impact: allocation, contracts, costs and spending, and performance verification (figure 1.1). Specifically, under the contracts domain, it recommended that the Global Fund could enhance the impact of its programs by incorporating explicit performance incentives into its grant agreements—that is, directly linking at least a portion of funding to incremental progress against the most important indicators.

Two years later, much has changed and improved under the Global Fund's New Funding Model. Nonetheless, many issues remain and challenges remain relevant: a stagnant medium-term funding outlook, donor demands for greater accountability, the challenges of low-quality data and verification of results, and the need to focus human and financial resources on the most effective interventions, in the most at-risk populations, in the most affected countries. New challenges have also emerged as pressing priorities: the quest for sustainability and transition away from donor funds in middle-income countries, the corresponding imperative for domestic resource mobilization, and expanding coverage goals that offer the promise of an AIDS-free generation⁶—but which, for now, remain aspirational within a context of limited resources and massive dropout along the HIV testing and treatment cascade.

Figure 1.1 Value for money domains for global health funders



Source: Glassman, Fan, and Over 2013.

In this context the Global Fund has expressed cautious enthusiasm for results-based grant designs as a potentially useful tool to accelerate impact. Free from year-to-year disbursement pressure (like government agencies) and rigid allocation policies (like the World Bank's International Development Association), the Global Fund is also uniquely equipped to push forward innovative financing models. But despite interest, the how of these next generation financing models remains a challenge. Realizing their potential requires technical know-how and careful, strategic decisionmaking that responds to specific country and epidemiological contexts—all with little evidence or experience to guide the way.

This report—the product of a co-chaired working group—thus addresses the how of next generation financing models—that is, the very concrete steps needed to change the basis of payment from expenses to something else: outputs, outcomes, or impact. Specifically, when and why is changing the basis of payment a good idea? What are the right things—for example, indicators and results—to

purchase from grantees? How much and how should grantees be remunerated for their achievements? How can the Global Fund verify that the basis of payment is sound—for example, that the reported results are accurate, reliable, and represent real progress against disease control goals? If discrepancies are found, how should it respond? What is needed to protect communities against coercion or other human rights abuses, ensuring that these new incentives do not drive unintended consequences? And at the end of this process, how can the Global Fund know whether the entire experiment is working with respect to its own objectives?

The working group did not just rely on the literature, previous CGD research, and working group deliberations to make its recommendations; it also convened a group of the best development and health economists in the world to bring a Nobel Prize–winning field of economics—contract theory and mechanism design—to the special challenges of aid financing for global health. The insights from this work—part of which is included in this report and associated background papers—have implications for the Global Fund and beyond.

Big ambitions, limited leverage and visibility: Why the principal-agent relationship and contract theory matter for global health financing

The Global Fund’s disease control goals are noble; they are also broadly shared by its grantees. Yet the structure of the traditional donor-recipient relationship may fail to fully align the incentives of both parties, compromising the speed and impact of Global Fund investments. This principal-agent problem is analogous to similar challenges faced in private sector contracting and the regulation of public sector monopolies—and the insights from those sectors suggest contracting mechanisms that can help get the incentives right.

In their diagnosis of the weaknesses of traditional aid models, Birdsall and Savedoff (2010) highlight three major challenges intrinsic to the donor-recipient relationship: weak accountability relationships, divergent goals, and asymmetric information. These issues—summarized and applied to the Global Fund below—collectively make up three dimensions of the principal-agent problem, which can lead to suboptimal disease control outcomes.ⁱ

i. Like other types of market failure, the principal-agent problem can, if not resolved, increase the cost or totally obstruct transactions between

First, accountability relationships can become muddled and fraught in donor-recipient interactions. Country governments should, in theory, be accountable to their own voting, taxpaying citizens. But when an external funder introduces donor money, the taxation-based accountability link between citizens and their governments can weaken.

Meanwhile, responding to their own citizens, “funders see themselves as responsible for ensuring that foreign aid is used appropriately. They therefore insist on determining standards...[that] reflect their own perspectives and requirements,” and often feel frustrated or cheated when the recipients fail to achieve the funder’s intended objectives.⁷ Funders attempt to regain accountability with controls on the use of funds, procurement, and data collection. But because donors have no independent authority in a sovereign country, the scope of punitive sanctions is highly constrained. At best, a funder could recoup stolen funds, cut funding for the next grant period, and publicize a country’s misdeeds to confer reputational damage; in contrast, it has no power to jail corrupt officials, fire incompetent bureaucrats, or vote the country’s leadership out of office.

Recipient countries also suffer from the lack of mutual accountability. For example, donors can unilaterally cut funding midway through a grant, with or without cause. Recipient countries may also resent donors’ micromanagement and inflexibility, plus the added administrative burden of complying with donor controls.

Second, while donors and recipients are aligned on disease control goals, each may hold divergent secondary objectives. The Global Fund (through the Country Coordinating Mechanism) enlists the Principal Recipient as its agent on behalf of its intended beneficiaries: the poor and sick in low- and middle-income countries.ⁱⁱ But Principal Recipients make imperfect agents because their secondary goals likely differ from those of the Global Fund, they have

parties, here the Global Fund and the Principal Recipient, that would be mutually advantageous and would also benefit third parties such as people in the recipient country at risk of HIV/AIDS, tuberculosis, or malaria. See Akerlof (1970).

ii. For each application, the Country Coordinating Mechanism nominates the Principal Recipient to implement the program; the nominated Principal Recipient is then subject to approval by the Global Fund following a capacity assessment. (In exceptional cases when a country portfolio is managed under the Additional Safeguards Policy, the Global Fund selects the Principal Recipient.)

fundamentally different resource constraints, and they can draw on local information that is unavailable to the Global Fund.

Finally, the Global Fund has less information than its recipients about the actual process and costs of implementation on the ground, an imbalance that economists call information asymmetry. Though the Global Fund is able to compile and apply technical expertise from the world's leading authorities on health technology and delivery, it is likely to know less than a country government about the local cost and effectiveness of different interventions or delivery strategies—and, importantly, the scope for potential cost savings or efficiency gains.⁸ Yet, following its current procedures, the Global Fund must negotiate with recipients regarding which inputs to buy, at what cost, and how those inputs should be delivered—all with relatively limited knowledge about whether those inputs appropriately reflect local market prices and whether they will be effectively transformed into better health services and population health. As a result, the Global Fund might overpay for certain inputs, or it might require the Principal Recipient to purchase an inefficient basket of goods that does not maximize health outcomes.

The Global Fund's current typical contractual relationship with a Principal Recipient thus specifies the transformation of money into inputs; the binding accountability relationship concerns whether those inputs are correctly purchased. The relationship between those inputs and health is separately measured and reported to the Global Fund, but with only an indirect contractual link to payment.ⁱⁱⁱ As a result, there are few explicit incentives embedded within the grant agreement to produce health services and health more efficiently with a given set of money and inputs.

Next generation financing models: A new approach to an old problem

The principal-agent problem arises because funders and recipients lack mutual accountability, have sometimes divergent objectives,

iii. Data may inform a holistic assessment of grant performance under the Global Fund's existing performance-based financing system; however, several factors limit the ability of performance-based financing to create a clear contractual relationship, including inadequate choice and number of indicators for contractual purposes; a methodology that includes absorption and expenditure rates among its considerations; and wide leeway for the Secretariat to adjust disbursements based on its judgment and contextual factors.

face different constraints, respond to different incentives, and have access to different information.

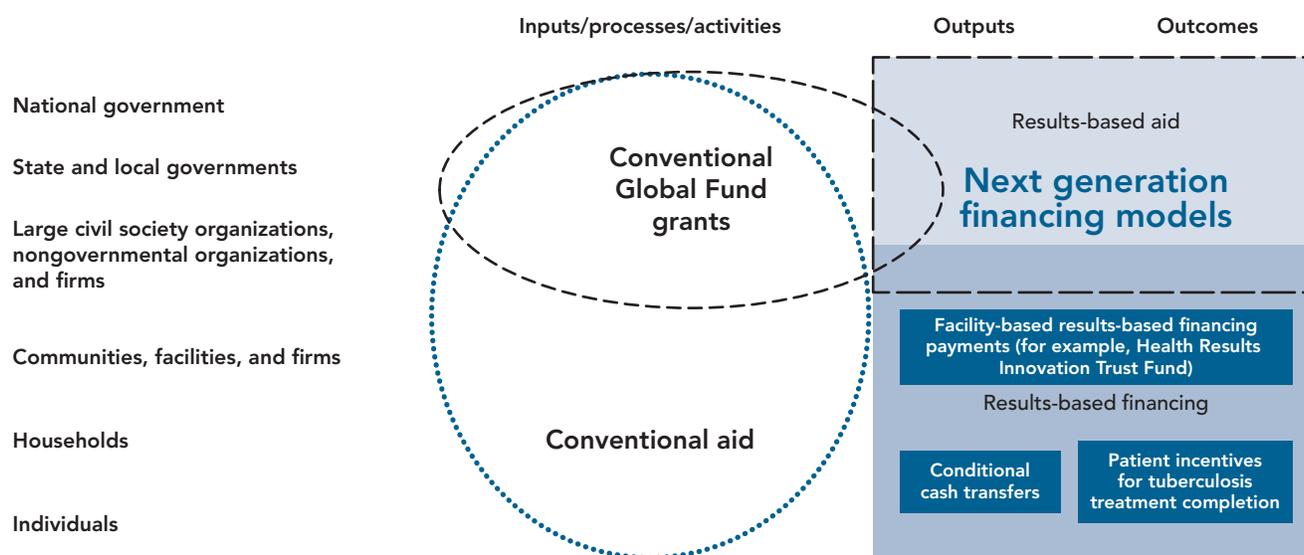
Next generation financing models—defined for the first time in this report as an accessible way to think about mechanism design in the aid relationship between a funder and a recipient—attempt to address these deficiencies by aligning incentives between donors and recipients. They create an explicit and enforceable contract between funder and recipient, wherein both parties agree that the financial relationship will be tied directly to achieving mutually important, realistic, and measurable gains in the provision of health services or population health.

These models fall within the broader universe of results-based financing and aid projects (figure 1.2). In contrast to facility-based results-based financing payments or household-based conditional cash transfers, next generation financing models frame the relationship between a global health funder and its grant or loan recipients (in Global Fund parlance, Principal Recipients). These recipients are the national or local governments and large civil society organizations that contract directly with the Global Fund. But like results-based financing and conditional cash transfers, next generation financing models focus on outputs or outcomes—not the inputs or processes that make up the basis of payment in conventional aid projects and grants. Next generation financing models also encompass several existing or proposed aid mechanisms—for example, Cash on Delivery Aid (national governments/outcomes), performance agreements to reduce deforestation (national or state/local governments/outcomes), and the World Bank Program for Results (national governments/outputs).⁹

Within next generation financing models, the contractual relationship might bypass inputs entirely; that is, the transformation of money into inputs and inputs into outputs or health would be outside the contractual scope of the agreement. Instead, contract negotiations and terms focus on appropriate remuneration and measurement for service delivery or health improvement. Recipients are accountable for delivering results that both they and the funder care about. Financial flows respond accordingly and predictably, given the results achieved.

These next generation grant designs can address several challenges of conventional grants. First, they establish a collaborative and explicit mutual accountability relationship between funder and recipient. Recipients are liberated from their biggest problems with donor aid: paternalistic micromanagement, extensive parallel

Figure 1.2 Next generation grantmaking within the results-based aid universe



Source: Adapted by authors from Perakis and Savedoff (2015).

documentation to fulfill donor requirements, and unilateral donor control over disbursement decisions. The funder also gains because it has a clear mechanism with which to hold recipients to account: it will only pay for the results that have actually been achieved.

Second, these agreements can align incentives for both parties toward their shared goal: health improvement. The recipient continues to hold alternative and potentially conflicting secondary objectives. However, recipients can pursue those objectives with donor funds only if they are able to efficiently achieve the donor's principal objectives and then apply cost savings to other priorities.

Finally, information asymmetries between the funder and Principal Recipient persist but become far less important. The Global Fund would still have an incomplete understanding of the cost structure and implementation arrangements on the ground—but because that information would no longer be tied to disbursements, the Global Fund would be less disadvantaged by those asymmetries.

A move toward next generation financing models requires a paradigm shift in grant administration, with major changes in both mindset and practice. At times, next generation grants will not

achieve results and thus not culminate in any or full disbursement—but this does not necessarily imply a grant management failure. Of course, a situation in which people at risk do not receive the health services they need is correctly considered a human and programmatic failure. But even those failed grants might represent effective grant management, because no donor funds were wasted on activities or governments that failed to deliver results. As a result, the Global Fund can redeploy those funds with new and better strategies, or via a different Principal Recipient with greater capacity to deliver results, in order to reach the people in need. This is in contrast to a conventional approach, where substantial donor resources might be spent, and the Principal Recipient still might not deliver results to the people in need.

At the same time, technical assistance becomes something demanded and requested by recipients to achieve shared objectives—not an external force imposed on them as a prerequisite for funding. Likewise, recipient countries' mindset and orientation also must adapt to accept and appropriately exploit the opportunities they receive as a result of enhanced responsibility and autonomy.

Chapter 2

Learning by doing: An evolving approach to incentives at the Global Fund and elsewhere

Since its creation, the Global Fund has aspired to link funding to results achieved, has established routine internal processes toward that end, and is one of the few donors to do so across its entire portfolio. Nonetheless, there is some perception that the Global Fund's original performance-based financing system has not fully succeeded in increasing programmatic performance, incentivizing innovation, or building sustainable country ownership. Adoption of next generation grant models, as described here, will thus require major shifts in the Global Fund's routine grant management procedures, including performance-based financing.

Performance-based financing: Ahead of its time, but diluted incentives and an uncertain future

Under the Global Fund's previous rounds-based funding system performance-based financing was a complex process fully integrated into the Global Fund's routine grant management. To trigger a new disbursement or apply for a Phase 2 grant renewal, Principal Recipients reported their own results against the grant's performance framework, typically comprising 15–20 process, output, and (infrequently measured) outcome indicators, alongside agreed-upon targets for each. The Local Fund Agents then performed a light-touch verification exercise, typically consisting of a desk review of data and a limited number of preannounced onsite spot checks. A percentage score was calculated for each indicator by dividing actual achievement by the target; those scores were then averaged across all the indicators on the performance framework, with additional weight given to "top 10" indicators. That aggregate performance score was translated into a letter grade and later adjusted and finalized by the Global Fund Secretariat. To complete the process, the grant score was translated into an indicative disbursement range, and the Global Fund Secretariat used the range to inform the final disbursement decision and amount. Actual disbursements were often outside the indicative

ranges generated by the performance-based financing system, for a variety of reasons.ⁱ

Several limitations of performance-based financing may have compromised its ability to motivate better performance from Principal Recipients. First, the complex, holistic, and discretionary process did not include a direct link between results and payments, lowering the predictability of consequences and payments for better performance and thus weakening the power of the incentive. Second, the grant score combined too many performance elements—including financial reporting and data quality, in addition to the dozen-plus indicators—many of which were not objectively measurable or directly relevant to the program's goals. Third, the system relied largely on grantees' self-reports, with only limited data verification and few consequences for discrepancies between reported and actual performance. Finally, some empirical research shows a weak link between the Global Fund's own performance metrics and its actual disbursements.¹⁰

Perhaps most important, performance-based financing operated by changing the allowable ceiling of grant expenditures—but not by changing the basis of payment. That is, grantees still had to incur and document verifiable expenses (for example, receipts) as the basis of allowable payment; good performance meant that they could incur more expenses, while bad performance would lead to a lower overall ceiling for expenses. But because allowable expenses remained the basis of payment, grantees were still required to comply with all financial, procurement, and governance procedures—which all had considerable transaction and overhead costs and together may have limited the scope for Principal Recipients to undertake creative problem solving and innovation to maximize their own impact.

i. Final disbursement amounts were often adjusted based on contextual factors—for example, procurement needs or absorptive capacity or to ensure continuation of essential services. Depending on the specific situation, the adjustment may or may not have been appropriate or necessary. See Fan and others (2013) for further detail and discussion.

In 2008 the first five-year evaluation of the Global Fund noted that the performance-based financing system created several problematic incentives. Performance-based financing “[had] evolved into a complex and burdensome system that [thus] far focused more on project inputs and outputs than on development objectives, departing from the vision of an outcome-based model.”¹¹ Evaluators also questioned the data used to inform disbursement decisions: “while the system generates extensive data, it often fails to provide the key elements of information required to inform judgments on effectiveness.”¹²

Under the New Funding Model, performance-based financing remains in place, albeit with some major changes. The elimination of the Phase 2 renewal, paired with the Country Team’s more active role in ensuring grant success, has greatly reduced, at least in theory, the likelihood that a grant will be canceled due to poor performance. Some within the Global Fund feel pressure to disburse countries’ allocations except in the case of blatant fraud or misuse of funds. Together, these forces suggest an overall weakening of performance-based financing incentives and call into question their continued relevance and utility in motivating greater effort or efficiency.

From pilots to strategy: The Global Fund explores new financing models

Over the last year or so, in a desire to drive faster impact and differentiate its engagement with countries and in recognition of some limitations of the current approach, the Global Fund has explored new financing models as potential tools to address many of its current challenges. Its goals for these next generation financing models include:

1. To achieve faster impact through increased coverage, higher quality, and better value for money.
2. To increase efficiency by reducing administrative and transaction costs, both for the Global Fund Secretariat and for country recipients.
3. To prepare countries for self-sufficiency and transition by increasing country ownership, promoting autonomy, and building political will to address neglected areas of the disease response.

To this end, an internal Global Fund working group has spearheaded a series of pilots in settings as diverse as Rwanda, the Solomon Islands, and Mesoamerica. The piloting process has indirectly mobilized disparate arms of the Global Fund around a common

mission, with participation and buy-in from many different constituencies. Perhaps most important, the Global Fund’s risk management and finance divisions have moved forward with piloting a new model of assurance that relies on independently verified programmatic results in lieu of detailed financial reporting and audit. This bottom-up piloting process has also led to several savvy design choices, reflecting serious and strategic thinking about the Global Fund’s role and leverage points and how they interact with recipients’ own incentive structures. Some examples include the use of small payments to be disbursed directly against evidence of progress in lieu of tying full grant amounts to performance targets, enhanced independent verification of data, and a focus on outcomes rather than inputs as the Global Fund’s ultimate strategic objective.

But learning while doing also implies learning from mistakes—and the piloting process has indeed led to greater clarity about the potential stumbling blocks for next generation grant designs, particularly around choice of indicators, how much to pay (without enough information on costs), verification methodology, and application of verification results to payment. The importance of contingency planning has also come into sharp relief as the Global Fund has grappled with its tolerance for nondisbursement in case of nonperformance. Finally, it has become clear that Country Teams are at times unprepared and underresourced to take on the demands of next generation financing grant design, with few technical resources within the Secretariat that can be called on to help. Getting the right people and expertise in place—and equipping Country Teams to call on that expertise when needed—will be a prerequisite for scaling next generation grant designs.

Though the application of next generation grant designs has remained limited to a handful of pilots for the time being, the Global Fund is using this experience to determine the extent to which they can be used to address portfolio, disease, or grant challenges. Expansion of the use of next generation grant designs, should it happen, would be part of a broader move toward differentiation based on strategic considerations and grant and country characteristics. To this end, the Global Fund is currently developing a payment for results framework.

Balancing promise against risk

The Global Fund is optimistic about the potential of new financing modalities to motivate faster progress against its three target

diseases. Yet that optimism is counterbalanced by recognition of real risks associated with these new approaches—especially because they are largely untested. Innovation implies flying blind, with few concrete experiences to inform design, avoid potential pitfalls, and reassure skeptics that the long-term benefits will outweigh the short-term costs. There will inevitably be a learning curve to navigate.

First, many stakeholders fear that explicit financial incentives might introduce unintended consequences, either creating greater inequities or driving unanticipated negative consequences. Principal Recipients could use unacceptable tactics to achieve faster progress against the indicators—for example, by coercing people who inject drugs to enroll in opioid substitution therapy. They could also game the system by trying to bump up their numbers while ignoring the spirit of the indicators. For example, Principal Recipients could try to achieve greater coverage of services by reducing the quality of their services; they could treat only the easiest cases, turning away marginalized populations or complicated cases (so-called “cream-skimming”); or they could exaggerate their reporting to claim a larger payment.

Others express concern about the programmatic risks of withholding disbursements. Many countries that receive Global Fund financing are low income, with few domestic resources to devote to the health sector. Other Principal Recipients are nongovernmental organizations, often with little cash reserves, that rely almost entirely on the Global Fund to support their operations. If a Principal Recipient fails to achieve its targets in the first year and thus receives a reduced disbursement, the Global Fund fears that the Principal Recipient will then have even fewer resources to achieve results in the next period—hence sparking a downward spiral of worsening performance. In addition, the Global Fund could face ethical quandaries or reputational damage if its decision to not disburse resulted in patients losing access to lifesaving health services.

Third, there is no guarantee that the benefits of a new approach will necessarily outweigh the costs. Especially in the short term, a move to next generation grants will require substantial time and money investment by both the Global Fund and Principal Recipients to retain appropriate expertise, communicate and negotiate with country stakeholders, and design and contract robust verification systems (though there may be long-term savings from reduced grant implementation and oversight costs). Those investments are justified to the extent that they result in substantially faster or improved progress against disease goals, but the evidence base on the effectiveness

of new modalities is still limited (see Part II), making it impossible to predict with certainty. Over time, careful evaluation can help quantify the costs and elucidate the benefits of the new approach.

Finally, some stakeholders are concerned about loosening the reins over Principal Recipients’ financial and programmatic management. Where systems and technical capacity are weak, Principal Recipients may expend considerable funds and effort pursuing technically inappropriate or mismanaged approaches, with little payoff for the disease response. Relatedly, there are reputational and organizational risks to the Global Fund and Principal Recipients associated with a new approach. The current risk framework—focused heavily though not exclusively on fiduciary responsibility—emerged in response to revelations of Principal Recipient misuse of some Global Fund money in 2011. Even if the Global Fund adds rigorous verification of activities, outputs, or results, it is possible that donors may still demand visibility into the end-use of Global Fund monies and reject the new assurance framework.

These risks can be managed and mitigated by savvy grant design (see Part II), but they cannot be eliminated. Yet there are also very real risks of maintaining the status quo: the risk of not achieving maximum impact, the risk of not having full visibility on the ultimate uses of funds, and the risk that ultimate goals may never be attained due to weak accountability relationships between the Global Fund and its grant recipients. The opportunity cost of the conventional approach may affect whether people in low- and middle-income countries continue to suffer the terrible costs of AIDS, tuberculosis, and malaria.

Looking outside: What can be gained by getting incentives right in global health?

Despite enthusiasm for results-based financing modalities, implementation has thus far been slow and cautious.¹³ The most extensive implementation and best evidence for results-based payments to health facilities in low- and middle-income countries has emerged from the World Bank’s Health Results Innovation Trust Fund. In Argentina,¹⁴ Rwanda,¹⁵ Zambia,¹⁶ and Zimbabwe,¹⁷ rigorous evaluations have found that better incentives can greatly improve health system performance via important gains in use and service quality—sometimes with measurable effects on health outcomes. But the success of facility-based payments has depended on smart design choices, careful attention to mitigate the risk of unintended

consequences, and a robust verification strategy. Where those elements were not in place—for example, in India and the Democratic Republic of the Congo¹⁸—attempted facility-level results-based financing programs have fallen flat.¹⁹

For other results-based aid modalities such as Cash on Delivery, experience is limited and evidence is scant. Just a handful of results-based aid programs have been implemented at scale, and most have not incorporated data collection on a counterfactual to enable a rigorous evaluation of their impact.²⁰

Even so, evidence abounds that incentives really matter for results—and the ubiquity of suboptimal incentives in global health financing is compromising our collective ability to maximize impact and contain costs. For example, in China school-based incentives led to a 14 percent drop in anemia prevalence—roughly the same effect as a large block grant but at about half the cost.²¹ In New Zealand the introduction of incentives for coverage of preventive services coincided

with a 30 percentage point increase in childhood vaccination and a 20 percent increase in screening for cardiovascular disease, all in just six years.²² In Germany financial incentives for better, coordinated patient care improved the effectiveness of diabetes management and substantially reduced three-year mortality and other complications, even as it saved money for each patient enrolled.²³ And in the United States a recent experiment evaluated the introduction of accountable care organizations to align provider incentives with those of patients and the payer (Medicare) for about 750,000 enrollees.ⁱⁱ The result was savings of \$385 million over two years for Medicare, all without detriment to patient experience or quality of care.²⁴

While there is also mixed evidence from high-income settings, the bottom line is that even if we do not yet know exactly what will work best—and even though the risks are real and considerable—getting the incentives right may be the single most effective strategy for stretching the impact of scarce global health dollars.

ii. Enrollment changed over the study period; $n = 675,712$ in 2012; $n = 806,258$ in 2013.

Part II

Nuts and bolts: A practical guide
to next generation grant design

Chapter 3

Changing the basis of payment: Why, where, and how much?

In pursuit of its mandate to prevent and treat HIV/AIDS, tuberculosis, and malaria, within the constraint of fixed replenishment resources, the Global Fund aims to support effective delivery of health services and facilitate countries' graduation to self-sufficient, effective disease control programs. To do so, the organization provides funding to grantees and contractors known as Principal Recipients. The Global Fund's grant portfolio is highly heterogeneous; grants vary substantially by size, Principal Recipient type, disease component, and number of subrecipients (figure 3.1).

The design of next generation grants is therefore intended to incentivize Principal Recipients to achieve faster disease impact and long-term sustainability. To help guide the way to better incentives, this chapter answers the following questions: What does it mean to change the basis of payment? Why might the Global Fund and its Principal Recipients want to change the basis of payment? And in what contexts would doing so be desirable—and where would doing so be feasible, either in full or in part?

Figure 3.1 The Global Fund grant portfolio at a glance (active grants as of mid-2015)

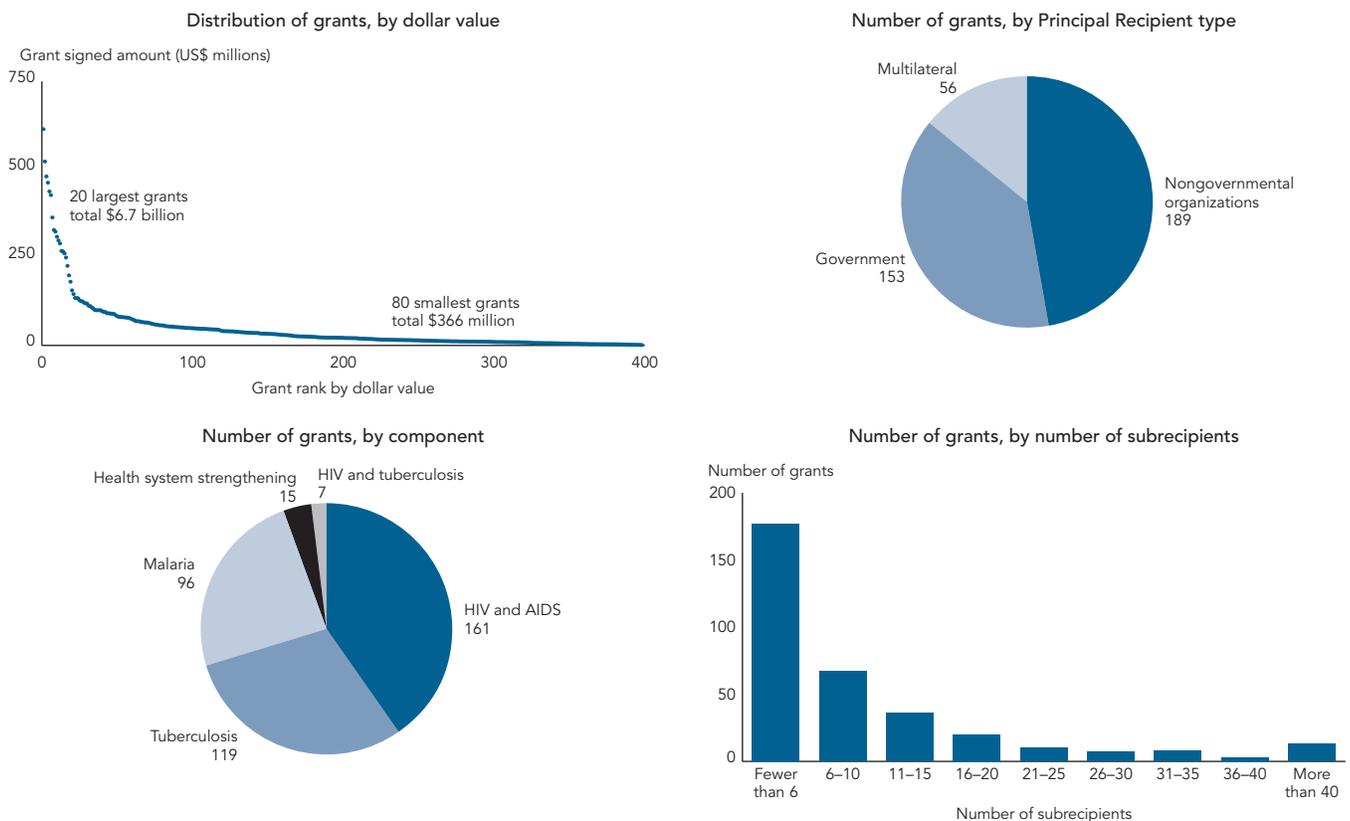


Table 3.1 Potential bases for payment along the results chain

RESULTS LEVEL	BASIS OF PAYMENT	REQUIRED EVIDENCE OR ASSURANCE	
1. Inputs	Expenses (for example, purchase of drugs or equipment or payment of salaries)	Submitted receipts, verified by audit	INPUT FINANCING
2. Activities	Payment for the implementation of prespecified eligible activities	Submitted receipts or activity records, verified by audit	
3. Outputs	Production of outputs	Submitted health service output records, verified by third party	PERFORMANCE-BASED PAYMENTS
4. Outcomes	Achievement of outcomes	Evidence supporting claimed change in behavior or clinical outcome from either facility data or population-based surveys ^a	
5. Impact	Achievement of impact	Evidence supporting claimed improvement in morbidity or mortality from household-based surveys	

a. The Cash on Delivery approach to foreign assistance advocates payment for outcomes and is thus one version of a performance payment.

What does it mean to change the basis of payment?

For the Global Fund and most other health donors, it is standard practice to use expenses as the basis of payment for grant agreements. That is, agreements specify allowable inputs to purchase with donor funds, and assurance and due diligence by the Global Fund focus on preapproving and verifying eligible expenses incurred. To continue receiving future disbursements and justify past expenses, grantees must collect, maintain, and share documentation with the Global Fund showing that funds were indeed spent on preapproved input and activities, that appropriate prices were paid, and that the appropriate fiduciary controls and procedures specified in the grant agreements were followed. If funds were spent on nonapproved inputs, or not spent at all, there is thus no valid basis for payment, and the Global Fund will withhold or attempt to claw back its disbursements.ⁱ

Expenses have thus been the traditional basis of payment for Global Fund grants, but they are not the only option. Indeed, the

basis of payment can in theory rest anywhere along the results chain developed as part of the literature on evaluation (table 3.1).²⁵ For example, the Global Fund could change the basis for payment such that it pays a fixed cost for specific service delivery packages (an output), for each individual completing a standard course of tuberculosis treatment and testing smear-negative (an outcome), or for declines in the disease burden as measured by a household survey (impact). In each case, expenses are no longer the basis of payment and thus no longer need to be preapproved, documented, or reported. Instead, assurance and due diligence shift to ensuring that the new basis of payment is sound—that is, that the Global Fund bases disbursement on accurate, verified information about the extent to which those goals have been achieved.

For the sake of simplicity, the remainder of this chapter groups different bases of payment along the results chain into two categories: input financing—the current, traditional basis of payment—and performance-based payments—which include outputs, outcomes, and impact and constitute next generation grant agreements.

Why might the Global Fund and its Principal Recipients want to change the basis of payment?

The Global Fund depends on Principal Recipients to deliver health services in recipient countries on the Global Fund's behalf. As described in Chapter 1, Principal Recipients can thus be considered

i. Grant agreements do prescribe a target number of processes, outputs, and outcomes that the grantee is expected to achieve with grant monies, and those results may result in changes to the ceiling of total disbursements under the performance-based financing system (described in Chapter 2). But regardless of whether the total amount of the disbursement changes, the basis of payment does not—payment is still based on whether Principal Recipients incur and document allowable expenses.

Table 3.2 Hypothesized benefits of changing the basis of payment

PERFORMANCE-BASED PAYMENTS	CONVENTIONAL INPUT FINANCING
Aligns incentives for speedier impact on mutually agreed outcomes.	Progress on mutually assured outcomes is not fully known due to limited performance verification.
Creates incentives to strengthen routine health information systems by providing regular checks on the accuracy of self-reported data.	Limited assessment of accuracy of performance data, coupled with input financing (but no incentives) to strengthen health management information system.
Global Fund staff freed to focus on whether grants are achieving agreed results.	Global Fund staff must focus on assuring that Principal Recipient adheres to pre-agreed inputs, budget, and activities.
Creates incentives for Principal Recipients and national governments to prefinance programs and, potentially, to commit domestic resources, thereby creating conditions for transition and long-term sustainability.	Only weak incentives (through the counterpart financing policy) for prefinancing or increased domestic investment in the disease response.

the agents of the Global Fund—where the Global Fund is the principal.ⁱⁱ

Within this principal-agent relationship, input financing models may fail to provide the Principal Recipient with sufficient incentive to improve efficiency by reducing costs or increasing output or outcomes, because the Principal Recipient may feel it has little to gain from those improvements. Because expenses are the basis of payment, every dollar the Principal Recipient spends (below the grant ceiling) will be covered by the Global Fund regardless of efficiency. Unexpended funds often may be accessed later, with the Global Fund's permission; but the Principal Recipient might face reduced allocations in future funding cycles (due to its lack of absorptive capacity). This characteristic of the grant—the degree to which the Principal Recipient directly benefits (or not) from cost savings—is referred to in contract theory as the power of the agreement. Since input financing models require Global Fund payments to be consumed by eligible mutually agreed expenses, nothing is left for the Principal Recipient to allocate to its own objectives. And the Principal Recipient has difficulty even

ii. Principal recipients not only are agents of the Global Fund; they also can be considered disease-specific agents of the Country Coordinating Mechanism, of the recipient government, and of the patients whom they serve. This report sets aside these other principal-agent relationships. This simplification can be justified either on practical considerations, because the Global Fund–Principal Recipient relationship is the immediate focus of study, or on theoretical grounds, based on the assumption that any divergence in the disease-specific interests of the Global Fund and recipient government and its population are artifacts of the poverty and short time horizons of the recipient government and its population.

reallocating expenditures from one eligible category of inputs to another. Thus, the pure input financing model has no power to motivate more efficient behavior from the Principal Recipient that aligns with the Global Fund's own objectives. By shifting the basis of payment to performance-based payments, the Global Fund can increase the power of the agreement and better align the Principal Recipient's incentives with its own strategic objectives. In so doing, it can theoretically drive greater impact, faster progress, and more efficient activities.

Beyond the power of the contract, its structure is hypothesized to drive several other benefits (see table 3.2 and Part I).

In what contexts would changing the basis of payment be desirable—and where would doing so be feasible, either in full or in part?

In theory, all Global Fund–Principal Recipient relationships face the same principal-agent problem, and thus shifting to other bases of payment is desirable. These kinds of agreements can work with Principal Recipients of differing capacities; strong capacity should not be considered a prerequisite for their consideration (box 3.1). But practical concerns may preclude the Global Fund from doing so in many settings, at least for all but a small portion of total funding. This section discusses four of the most important considerations for where and to what extent the Global Fund can change the basis of payment: whether the goals of an agreement are contractible, whether the basis of payment can be measured and verified with sufficient precision to inform payment, whether the risk of nonpayment is acceptable to the Global Fund, and whether the agreement is sufficiently acceptable to the Principal Recipient such that it agrees to participate.

Box 3.1 Limited-capacity Principal Recipients and next generation financing models

In considering the feasibility of next generation financing models, the capacity of Principal Recipients is frequently expressed as a concern. Many believe that low-capacity Principal Recipients need intensive Global Fund oversight and targeted investment to succeed; without that, skeptics suggest that the Global Fund could be setting such Principal Recipients up for failure.

As expressed elsewhere in this chapter, there are indeed relevant capacity constraints that may affect the Global Fund's choice of grant model. Most essentially, the results data that serve as the basis of payment must be ironclad. Independent verification of results is essential for Principal Recipients of all capacity levels (see Chapter 5)—but if payment is to be based on clinical or routine programmatic data, the Principal Recipient's health management information system must produce reasonably complete and accurate self-reports that can later be verified by the independent third party. To the extent that Principal Recipients may need to prefinance a portion of service delivery costs, illiquid Principal Recipients will be ineligible for next generation grant models.

However, there is strong reason to believe that next generation financing models can work well with Principal Recipients of differing capacities, at least among those that meet preexisting Global Fund capacity requirements. (We assume that a minimum level of programmatic capacity is already ensured by routine due diligence, whereby the Country Coordinating Mechanism nominates the Principal Recipient and the Global Fund confirms that selection following a capacity assessment.)

First, there is no inherent reason to believe that low-capacity Principal Recipients will be any less able to deliver

services under a next generation financing model arrangement than they would under traditional input financing; indeed, they should be more motivated to improve management and effective delivery practices.

Second, some capacity issues are actually obviated within next generation grants. For example, indigenous nongovernmental organizations may be highly effective at delivering services to their communities, yet struggle to comply with complex Global Fund financial and procurement requirements. Under a next generation grant, those capacity constraints become less relevant, allowing the nongovernmental organization to focus on delivery of results instead of administrative recordkeeping.

Finally, Principal Recipients maintain their ability to request and receive technical assistance from the technical partners—and given the new incentive structure, they should be highly motivated to request the right assistance to achieve faster progress against the mutually agreed indicators. Global Fund Country Teams can play a facilitating role with this process, helping connect country stakeholders to the appropriate sources of expertise. Country teams can also serve as partners within the planning process, helping Principal Recipients to map the ingredients for success. And when a mutually desirable contract design requires that an illiquid Principal Recipient prefinance a portion of service delivery costs, the Global Fund could help the Principal Recipient obtain that prefinancing from the recipient government, perhaps via a World Bank loan or a Development Impact Bond.¹

Note

1. Center for Global Development 2013.

Are the goals of the grant contractible?

Changing the basis of payment is only feasible where program goals can be captured by one or more contractible indicators.²⁶ A contractible indicator is one that meets the following criteria:

Required

- It must be at least partially influenced or under the control of the Principal Recipient or health system and amenable to change within the timeframe of grant implementation.²⁷

- It must be measurable.
 - The measurement of actual achievement should be objective and direct—for example, not derived from self-reported behavior²⁸ or modeling.ⁱⁱⁱ
 - The measurement should be highly precise, valid, and reliable.
 - The measurement of the indicator should be affordable.

iii. While the actual results must be observed, the payment could be for improvements against a modeled trajectory. See Over (2011) for examples.

Table 3.3 Sample contractible indicators

	INDICATOR	WHY CONTRACTIBLE?
HIV	Number of people living with HIV who initiated antiretroviral therapy, with an undetectable viral load at 12 months	<ul style="list-style-type: none"> • Directly influenced by health system. • Direct proxy for a meaningful and important health outcome. • Measurable, and increasingly measured as part of routine care. • Independently verifiable. • Captures success of entire testing and treatment cascade.
	HIV incidence ^a	<ul style="list-style-type: none"> • Directly influenced by health system. • Direct measurement of a meaningful and important health impact. • Measurable and independently verifiable in some settings given investment in large-scale seroprevalence surveys.
Tuberculosis	Number of tuberculosis patients with drug susceptibility testing at diagnosis (in high-resistance areas only)	<ul style="list-style-type: none"> • Directly influenced by health system. • On causal chain to a meaningful and important health outcome. • Measurable, and measured as part of routine care. • Independently verifiable.
Malaria	Number of women attending antenatal clinics who received three or more doses of intermittent preventive treatment for malaria	<ul style="list-style-type: none"> • Directly influenced by health system. • On causal chain to a meaningful and important health outcome. • Measurable, and measured as part of routine care. • Independently verifiable.
Health system strengthening	Percentage of health workers who correctly follow clinical guidelines for management of specified condition	<ul style="list-style-type: none"> • Directly influenced by health system. • On causal chain to a meaningful and important health outcome. • Measurable and independently verifiable through direct observation.

a. Hallett and Over (2010) propose that the incidence estimate derived from successive HIV prevalence surveys is a contractible indicator of HIV prevention success in situations where baseline incidence is at least 1 percent per year, the anticipated decline in incidence is as large as half a percentage point, and the payment period is two years, provided the sample size is at least 50,000 in each community subject to the reward.

- When population measurement of an indicator is unaffordable or infeasible, indicators should be estimated with large and scientifically designed samples.^{iv}
- It must be independently verifiable.
 - The independent verification should be highly precise and reliable.
 - The independent verification should be objective and direct—for example, not derived from self-reported behavior²⁹ or modeling.
 - The independent verification should be representative.
 - The independent verification should be affordable.
- It must be a direct proxy, based on robust evidence, for a meaningful and important health outcome, or it must be on the direct causal chain, based on robust evidence, to a meaningful and important health outcome.
- It must not incentivize perverse outcomes, coercion, or human rights abuses.

Desired

iv. The sample design would depend on both epidemiological context and the payout schedule—since the sample size will determine the width of the confidence interval. See the discussion below on payout schedules.

- It should be difficult or impossible to manipulate or game without actually achieving the intended health outcome.

Table 3.4 Sample noncontractible indicators

	INDICATOR	WHY NONCONTRACTIBLE?
HIV	Percentage of men reporting use of a condom the last time they had anal sex with a male partner	Not directly measurable or verifiable; based on unreliable self-reported behavior.
	Number of condoms distributed	Without information on proper use or number and type of beneficiaries, it is not on direct causal chain to a meaningful and important health outcome.
Tuberculosis	Tuberculosis mortality rate	Current measurement in most settings is based on modeling, not direct observation.
	Treatment success rate if delinked from case detection	A higher treatment success rate could indicate either better quality of care and treatment outcomes (a good thing) or lower case detection or cherry picking of the most easily treatable cases (a bad thing). Could reward Principal Recipients for reducing the number of people treated or turning away complicated cases or hard to reach populations.
Malaria	Number of confirmed malaria cases	Could indicate either enhanced case detection and diagnosis (a good thing) or higher malaria incidence (a bad thing). If the goal is to decrease the number of malaria cases, Principal Recipients can game the measurement by failing to conduct or record proper diagnostic procedures or to register the case in the surveillance system. Robust verification is not possible because of the impossibility of verifying a negative.
	Number of long-lasting insecticide-treated nets distributed through mass campaigns or continuous distribution	Without information on proper use or number and type of beneficiaries, it is not on direct causal chain to a meaningful and important health outcome.
Health system strengthening	Number of trainings held	Without information on whether training leads to better service delivery, it is not on the direct causal chain to a meaningful and important health outcome.

- Where possible, a single indicator that accounts for an entire testing and treatment cascade is preferable to separate indicators for each step along that cascade.^v
- It should cover a meaningful or important part of the disease response in a given country; where the disease response has multiple essential pillars, it should not incentivize just one pillar at the expense of other essential elements.
- Ideally, it should be routinely measured through the national data system.
- It should not obstruct existing health systems approaches within the country—for example, it should promote integration and not undermine overall health system functioning.

What does this mean in practice? Tables 3.3 and 3.4 provide examples of contractible and noncontractible indicators for each of the Global Fund's three target diseases; these are indicative examples, not a comprehensive list.^{vi} Not all program goals will necessarily correspond to contractible indicators. Malaria programs present a particular challenge (box 3.2). Where no contractible indicator can be found, programs should be funded through traditional input-based grant mechanisms.

Importantly, an indicator that is contractible in one setting will not necessarily be so in another. The choice of indicator is a strategic decision that should be based on the specific country context—for example, the problem to be solved, the population to be reached, and

v. For a definition and illustration of the AIDS treatment cascade, see <http://www.cgdev.org/blog/question-quality-why-retention-matters-aids-treatment>.

vi. Health systems strengthening is, by its nature, less contractible and thus a lower priority for the inclusion of results-based financing features.

Box 3.2 Contractible indicators for malaria impact: A unique challenge

As outlined in Part II, identification of contractible indicators is a prerequisite for designing and implementing next generation grant models. This requirement presents a particular challenge for malaria. Malaria grant managers are enthusiastic about the potential to provide programmatic flexibility and a simplified grant approach but have struggled to find a contractible impact indicator for malaria that could serve as the basis of payment. Two basic approaches have been proposed—each with significant problems.

The first approach would rely on surveillance systems to report the number of confirmed malaria cases; the Global Fund would then pay the Principal Recipient for reducing the number of confirmed cases over time. This approach has three major sets of problems:

1. Changes in the measured number of confirmed cases may bear little relationship to changes in the true number of clinical malaria cases. Diagnosis of a confirmed malaria case requires parasitological diagnosis, most frequently and easily established by a malaria rapid diagnostic test. Yet standards for reporting may vary greatly in different settings or be in flux over time; for example, before the advent of rapid diagnostic tests, many countries would presumptively diagnose all or most fever cases as malaria. The net result is that many different factors—most unrelated to the true burden of malaria—can change the measured value for number of confirmed cases. For example, rapid rollout of rapid diagnostic tests could either increase the measured value (if it resulted in confirmation of previously unconfirmed cases) or decrease the measured value (if it resulted in re-classifying presumptively diagnosed cases as nonmalarial fevers). Likewise, introducing community health workers to rural areas can expand access to malaria diagnosis and treatment services, saving lives—but also increasing the number of malaria cases detected and reported.
2. There is no simple way to independently verify the Principal Recipient's self-report for number of confirmed malaria cases. Indeed, the Global Fund could even perversely reward a country for failing to diagnose, confirm, or report malaria cases to the central surveillance system, because there is no easy way for a third party to verify whether the surveillance system had accounted for all cases.
3. The underlying true value is often outside the full control of the health sector. Emerging insecticide resistance can

lead to an uptick in malaria cases, even where a national malaria program has maintained high coverage of the best available vector control interventions. Year-to-year and seasonal climactic variation can also increase or decrease the burden over time. Finally, malaria may rise or fall over time due to secular trends—for example, increased wealth and urbanization—without any action by the health sector.

Thus, number of confirmed malaria cases fails as a contractible indicator against at least three of the contractibility criteria: its measurability, its independent verifiability, and the health sector's control over its true value.

For high-burden countries, a second approach would use representative household surveys to measure parasite prevalence within a population at risk—and the Global Fund would pay the country for lowering its parasite prevalence over time. This indicator solves the independent verifiability problem noted above; the surveys themselves would be conducted by an independent third party. But this indicator likewise falls short for two main reasons.

- First, parasite prevalence is an insufficiently granular or precise indicator for contractual purposes. Differences of a few percentage points are largely meaningless; only orders of magnitude (50 percent versus 10 percent versus 1 percent) provide a real indication of improving health status. But those changes become clear only over intervals of five or more years—too long a time horizon given the Global Fund's three-year grant cycle. In addition, typical surveys are limited to a 5 percentage point or more margin of error, making them insufficiently precise instruments for payment.
- Second, as with number of confirmed cases, parasite prevalence may not be under the full control of the health sector.

The bottom line: there are many opportunities to use results-based approaches for malaria programs, but there is no single impact-level indicator that can serve as the basis of payment. Opportunities to adopt next generation financing models for malaria must thus be country-specific and at an earlier stage in the results chain. The specific choice of indicators—and whether any given output or outcome indicator is contractible—will depend on the epidemiological context, proposed activities, and local challenges in any given grant or country. (See, for example, the malaria case study laid out in Chapter 6.)

the objectives against which the Global Fund and country would like to see greater, faster progress. There is no one-size-fits-all indicator that will work perfectly in all countries at all times; nonetheless, thinking through the above criteria can help determine whether a candidate indicator is appropriate, measurable, and fit for purpose to inform disbursement decisions.

Finally, most contractible indicators are numerical. Some will capture elements of quality; for example, antiretroviral therapy retention and viral load suppression can only be achieved through the delivery of high-quality care. In other cases, however, explicit adjustments may be needed to account for quality of care or equity in service delivery. For example, a quality adjustment factor, based on measurable aspects of service delivery and patient satisfaction, could be applied to scale the quantitative indicator of achievement.

Are those contractible indicators measurable and verifiable in that context?

The Global Fund must have confidence in its ability to measure and verify the output, outcome, or impact indicators that are to become the basis of payment. And just as the Global Fund must now do due diligence to ensure the strength and oversight of Principal Recipient fiscal management when advancing money, so too must it exercise due diligence to ensure the strength of the measurement systems on which it will rely before signing on to a performance-payment model.

In short, there are two basic options for how the measurement could be conducted. The first is to rely on Principal Recipient self-reports based on internal health management information systems and then to verify those reports using an independent verification agent (verification is discussed in more detail below). As a prerequisite for the use of this approach, the Principal Recipient must have the capacity to produce initial self-reports of sufficient quality that they can be verified. That is, they must be reasonably complete, reasonably accurate, and clearly correspond to electronic or paper-based records of individual beneficiaries at the facility or community level. If the health management information system is not up to this standard in a specific country or setting, Principal Recipient self-reported indicators cannot be used as the basis of payment. However, the Global Fund can instead pay the Principal Recipient for complete and accurate data and reporting as a preparatory stage for a later shift to payment for results.

The second approach is for the Global Fund to contract independent measurement or piggyback on other independent measurement exercises such as Demographic and Health Surveys or representative population-based or facility surveys that do not rely on Principal Recipient self-reporting. Under this approach it becomes impossible to pay for incremental counts of patients or beneficiaries receiving services, and it is difficult to pay for clinical outcomes that would rely on accurate facility records. But there are many advantages: the Global Fund may be able to free-ride on preexisting measurement exercises for its own purposes, large surveys can detect impact-level changes like disease prevalence and morbidity, and independent measurement can be conducted anywhere, regardless of the Principal Recipient's monitoring and evaluation strength or capacity.

Is withholding payments acceptable to the Global Fund?

The power of the incentives created by next generation financing models is directly related to the amount of money at risk—that is, payments that take place only if the Principal Recipient achieves specified results. In contract theory, this requirement is referred to as the Global Fund's power to commit and means the Global Fund has the ability to—and will—withhold results payments when they are not earned. In many cases this basic premise presents no conflict. In other cases the risk of nonpayment may be ethically unacceptable for the Global Fund, or it may fail to serve the Global Fund's strategic interests.

First, the Global Fund considers itself ethically obligated to ensure continuity of essential services in recipient countries. Essential services can be broadly defined to include all lifesaving treatment or preventive service delivery—a huge portion of the Global Fund's overall portfolio. In many countries the Global Fund contribution finances primarily a steady supply of drugs and other commodities; to cut off funds could thus prevent patients from accessing lifesaving medicines. The Global Fund would not want to cut off access to essential commodities or services in the case of poor performance; thus, those commodities and services should be placed outside the scope of next generation financing models.

Second, in some cases the risk of nonpayment may not always serve the Global Fund's strategic interests. For example, in post-conflict settings the Global Fund might believe that a reinjection of capital is needed to rebuild infrastructure and reestablish

high-quality health services—perhaps not immediately, but over the course of several years. In such a case withholding funds might be counterproductive, even if performance is poor in the short term.

Third, global health organizations have traditionally been judged by donors on their ability to disburse funds—and the implications of that overall imperative can trickle down to the staff level. Global Fund staff may thus feel pressure to maintain a quick and high disbursement rate. Further, the Global Fund typically advances financing to the Principal Recipient for some period of service delivery—meaning that staff members might have to either claw back that money from the Principal Recipient in the case of poor performance or deduct from the next disbursement. Designing staff evaluation systems that change these incentives presents a major challenge.^{vii}

Is the agreement acceptable to the Principal Recipient?

Because the Global Fund relies on Principal Recipients to deliver services—and because the Global Fund is currently constrained in its choice of Principal Recipient^{viii}—grant agreements that change the basis of payment need to be acceptable to both parties. In contract theory, this is referred to as the participation constraint—essentially meaning that for a mechanism to work, it must be acceptable to the agent.

Why might changing the basis of payment be unacceptable to a given Principal Recipient? First, the costs of service delivery are hard to predict in advance and may change substantially when drug, salary, or gasoline prices rise or when service delivery is more costly than anticipated. Thus, relative to input financing, Principal Recipients might assume the risk for costs that turn out higher than expected. Second, Principal Recipients may not feel that achievement of the indicators is fully within their control—another source of risk. And even if achievement of those indicators is within their control, Principal Recipients may feel that they need initial startup investments in order to succeed—and they may be unwilling or

unable to put up their own funds to prefinance such investments for the first period of service delivery (see box 3.3).

Indeed, the current inability of Principal Recipients to prefinance service delivery is likely a constraint on efforts to change the basis of payment. Some Principal Recipients may have little cash reserve and lack access to credit markets. This lack of liquidity makes it difficult to deliver health services without donor support or to prefinance health service delivery in the expectation of future ex post payments for results.

The Global Fund must thus design grant agreements that incentivize faster progress while remaining attractive to and feasible for the Principal Recipients themselves. Prefinancing specifically is generally easier for the Principal Recipient if the grant is relatively small or if the proportion prefinaanced is small. The ability to prefinance also depends on the nature of the Principal Recipient—for example, a government Principal Recipient has access to government financing, particularly in middle-income countries, as well as many reasons to prefer prefinancing—for example, to reduce unpredictability of funds flows to essential health interventions or to eliminate the time pressure around the disbursement process. Finally, the Global Fund might facilitate access by a nongovernmental organization to prefinancing from a private sector social investor, as envisaged in the Development Impact Bond mechanism.³⁰ And Principal Recipient prefinancing may also offer important benefits for sustainability in addition to its enabling role in next generation grant agreements (box 3.3).

More generally, to increase the appeal of different bases of payment, the Global Fund could let the Principal Recipient keep a greater portion or all of the savings; increase the overall funding allocation (ceiling) for Principal Recipients that accept the additional risk; relax other fiduciary or oversight requirements, thus saving the Principal Recipient from administrative costs and paperwork; or design new financing mechanisms that also insulate the Principal Recipient from downside risk, as discussed in the next chapter.

A modular approach to changing the basis of payment

The decision to change the basis of payment need not be all or nothing; instead, grant design can lie anywhere along a continuum from 100 percent input financing to 100 percent payment for results

vii. Over and Ravallion (2015) propose a mechanism that could be used to reorient staff rewards at the World Bank away from disbursement and toward country performance.

viii. Principal Recipients are nominated by the Country Coordinating Mechanisms and approved or rejected by the Global Fund following a capacity assessment, in contrast to a standard competitive tendering process.

Box 3.3 Increasing Principal Recipient prefinancing: A good idea for the Global Fund

In its current model the Global Fund typically advances grant financing to the Principal Recipient for budgeted activities; it later requires the Principal Recipient to submit a record of its expenses as a condition for future payments. If the Principal Recipient has spent less than the full amount or has spent part of it on ineligible items, the Global Fund can subtract the unspent or unjustified amount from a future advance or can attempt to recover it. The latter action—sometimes referred to as claw back—is notoriously difficult, especially when a sovereign government is involved.

As an alternative, the Global Fund could require the Principal Recipient to prefinance all or some portion of expenditures on grant-related activities and subsequently request reimbursement from the Global Fund. From the Global Fund’s perspective, this can facilitate a move toward

next generation grant agreements—but it also offers several other advantages.

First, reimbursement for unjustified expenditures can simply be withheld, avoiding the need to reduce the amount of a subsequent advance or to claw back the funds and reducing pressure on the Global Fund staff to pay out more than the results agreement would justify. Second, prefinancing shifts a portion of the financial risk for poor grant performance away from the Global Fund and to the Principal Recipient, arguably enhancing country ownership of the program. Finally—and most important—prefinancing requires the Principal Recipient to assume more financial and managerial responsibility for health service delivery up front—for example, establishing budget lines for the activity—and thus can strengthen sustainable management, budgeting, and financing systems.

(figure 3.2). Any input-based grant can introduce at least a modest payment for results. The degree and extent to which payment for results is appropriate and strategically sound will vary substantially according to context.

Figure 3.2 illustrates four hypothetical grant models along the continuum, each of which adopts a different modular combination of input financing and result payments based on local context, constraints, and strategic objectives. Grant A, a large and complicated grant to a low-income government Principal

Recipient, would be closest to the Global Fund’s traditional input-based grant model. In such a setting, the disease program would depend largely on Global Fund contributions and cover many different activities. Thus the Global Fund would be reluctant to introduce the possibility that implementation problems could prevent disbursement of a large portion of its funds. Nonetheless, the Global Fund could use a relatively small supplemental payment to motivate greater effort and faster progress against the most important indicators.

Figure 3.2 Grant designs along the input financing to payment for results continuum

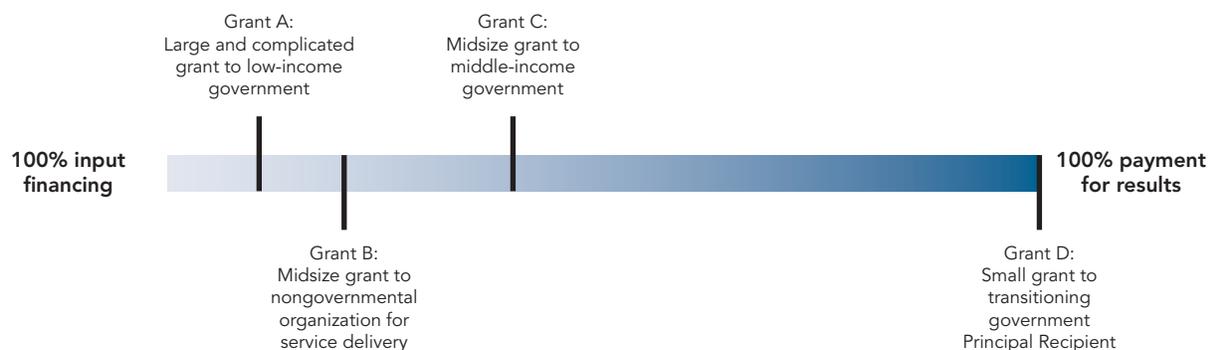
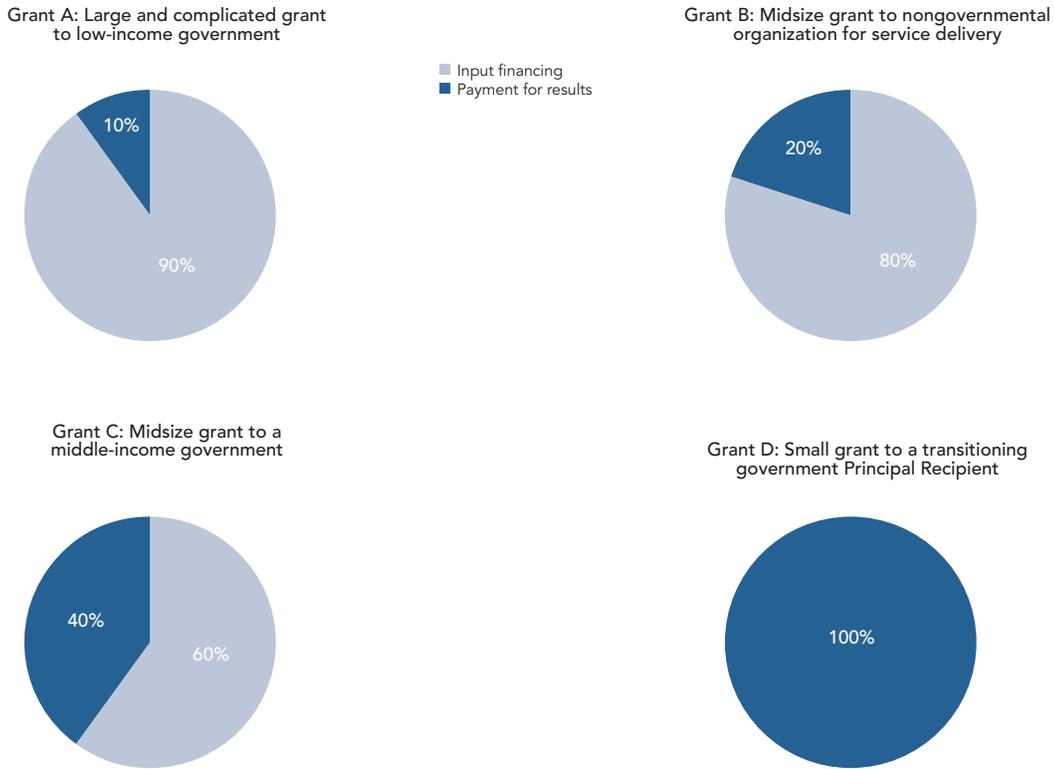


Figure 3.3 Four hypothetical grant models


Grant B represents a midsize grant to a nongovernmental organization to fund service delivery—for example, to deliver HIV prevention, testing, and counseling services to men who have sex with men. Because the Principal Recipient is a nongovernmental organization, it is unlikely to have sufficient liquidity to prefinance service delivery; it will also rely on the Global Fund to deliver essential commodities such as condoms. The Global Fund would be reluctant to risk discontinuation of those services in the case of slower than anticipated progress—but it would also want to quickly improve testing in the target population and linking individuals living with HIV to care. Thus a moderate portion of the grant—say, about 20 percent—could be used to supplement input financing and incentivize greater effort in serving those communities.

Grant C represents a midsize grant to a middle-income country, where a substantial portion of the Global Fund’s overall contribution covers commodity purchases and the remainder covers

miscellaneous elements of service delivery. In such settings the government could be reasonably expected to prefinance service delivery with domestic resources—and partial nonpayment would not present a serious financial challenge for the country government. Thus the Global Fund might guarantee funding for commodities but tie the remainder of the allocation to performance against strategic objectives.

Finally, Grant D represents a small grant to a transitioning government Principal Recipient—for example, a government that will lose access to Global Fund financing after this grant cycle. Here, it is reasonable to expect the Principal Recipient to prefinance all or a substantial portion of the grant. These grants are good candidates for changing the basis of payment for all or a large portion of funds.

Figure 3.3 provides another view of the four hypothetical grant agreements, clearly illustrating the scope for different modular combinations between input financing and payments for results.

Chapter 4

Getting the incentives right: Next generation financing models

Drawing from contract theory and mechanism design, this chapter describes next generation financing models that can help align the Principal Recipient's incentives with those global HIV/AIDS, malaria, and tuberculosis control goals that are embodied in the mission of the Global Fund.³¹ Each has different advantages, disadvantages, and prerequisites; the right strategic choice must thus be informed by the country context, particular disease control objectives, and Principal Recipient characteristics. The options described below do not give the full universe of all possible results payment designs but simply offer several potential starting points.

Starting simple: Fixed price models

In fixed price mechanisms the Principal Recipient is paid a fixed unit price for every unit of verified performance or improvement, with possible adjustments for quality. Fixed price is a large umbrella category of contract designs; aid instruments like Cash on Delivery and Verified Service Delivery fall under its auspices, as do many other potential designs.

Besides their simplicity, fixed price models offer many important advantages. They hold strong incentive power for Principal Recipients to improve their output and efficiency because the Principal Recipient can keep the savings—and, provided the payment per unit is higher than the Principal Recipient's cost of producing an additional unit, each additional person reached increases that surplus. Fixed price can also be applied to multiple outputs or outcomes, with a separate price for each. And compared to traditional models, the grant-related transaction costs of Principal Recipients are far lower; they no longer need to track receipts or comply with burdensome regulations but instead simply need to ensure that they achieve the intended results. However, fixed price models can be risky to the Principal Recipient because per unit payments are fixed, even if the Principal Recipient's costs are unexpectedly high due to factors outside of its control.

How to pay: Determining the payout schedule

A payout schedule specifies how different levels of Principal Recipient achievement relate to different levels of payment. Under the fixed price umbrella, the payout schedule can take a variety of forms (see boxes 4.1 and 4.2 for two real world examples). This section details four essential considerations for determining the payout schedule: the purpose of the fixed price, whether the fixed price rewards all performance or just improvements over a baseline, whether payments are incremental or in target-based tranches, and whether unit payment changes as achievement increases.

- *What is the intended purpose of the fixed price payment?* Fixed price payments can serve three basic roles—and the choice of role will influence the payout schedule. First, fixed price payments can serve as a supplement on top of input financing—either from the Global Fund itself, the government, or another donor. Supplemental payments can be operationalized in almost any grant, even if most input costs are paid by another donor such as the U.S. President's Emergency Plan for AIDS Relief. They work by incentivizing the Principal Recipient to efficiently translate existing inputs into better health results.

Second, fixed price payments can subsidize achievement of results—that is, fund some portion (but less than 100 percent) of the actual cost. Subsidy fixed price payments work best in situations where the government is willing and able to commit some resources to the disease programs but the Global Fund wants to encourage the government to both expand coverage and increase its share of total financing. For example, the Global Fund could offer to pay \$100 for each patient enrolled and retained on antiretroviral therapy—less than the actual (and often unknown) cost of service delivery—with the expectation that the country government will cofinance the remainder of the per unit cost. This structure would incentivize increased coverage of an intervention while eliciting greater domestic resource mobilization and encouraging the government to find efficiency

Box 4.1 Real world example: The Salud Mesoamerica Initiative

The Salud Mesoamerica Initiative is a five-year, \$114 million regional initiative to reduce maternal and child health inequities in its eight target countries. It is funded jointly by the Bill & Melinda Gates Foundation, the Carlos Slim Health Institute, the Government of Spain, and the Inter-American Development Bank.

Under the initiative, countries receive an investment tranche from the Inter-American Development Bank to support goals in target areas; countries must match that investment tranche with an equal sum of counterpart funding. Later, countries can receive a prize worth half their counterpart funding if they achieve sufficient progress against the initiative's core indicators. Receipt of the results payment is delivered on an all-or-nothing basis; to get it, a country needs to hit at least 80 percent of its targets across 9–12 indicators—meaning that it can fall short on one to two indicators at most.

Source: Ibarrarán 2015.

improvements. The Global Fund could even decrease its subsidy over time to create a clear path toward sustainability. Like supplemental payments, subsidy payments can work when Global Fund resources are supplementing those from other donors.

Third, fixed price payments can substitute for input financing—that is, fund the entire cost of service delivery. Here, the fixed price payments should closely mimic the actual cost of service delivery but allow the Principal Recipient to keep at least part of the savings where efficiencies can be achieved. They work by encouraging the Principal Recipient to find a more efficient way of delivering the service and then delivering the service to more people, allowing the Principal Recipient to keep the savings. Substitute fixed price payments are most feasible in contexts where the Global Fund pays for the entire cost of service delivery (that is, there are no other donors funding a portion of production), where the true cost of service delivery is already known (see Part III for advice on addressing other scenarios), and where the Principal Recipient is able to prefinance service delivery (limiting the extent to which the Global Fund might need to claw back advances).

- *What does the fixed price cover—just improvement, or everything?* The Global Fund can choose between two basic options for the

Box 4.2 Real world example: Cash on Delivery for education in Ethiopia

In 2012 the United Kingdom's Department for International Development partnered with the Government of Ethiopia to improve secondary education through a Cash on Delivery approach. Supplementing its preexisting general support to the education sector, the department agreed to pay an additional sum to the Ethiopian government for each extra student (on top of baseline levels) who completed primary school and took the final exam. It also paid an equal sum for every additional student who passed the final exam, up to a combined ceiling of £10 million each year. To encourage equity, it made a larger per student payment for girls and for children living in more disadvantaged regions. The department contracted an independent verification entity to analyze the government's reported results and verify their accuracy by visiting a sample of schools.

Source: Perakis and Savedoff 2015.

scope of its payment option. First, it could offer a fixed price payment for achievement above the Principal Recipient's baseline level.¹ This option is most appropriate for supplemental payments: since the supplement is intended to incentivize better health results with a given set of inputs (for example, to reward improved efficiency and effort), the Global Fund should pay only for improvements above the expected level of performance. Second, the Global Fund could tie fixed price payments to the entire output of the Principal Recipient—baseline levels included. This option is most appropriate for subsidy or substitute payments; here, the Global Fund is offering fixed price payments instead of input financing, so it would need to apply them even to maintain baseline levels.

- *Is the payout schedule continuous or lumped into tranches?* To create a simple payout schedule, the Global Fund could define

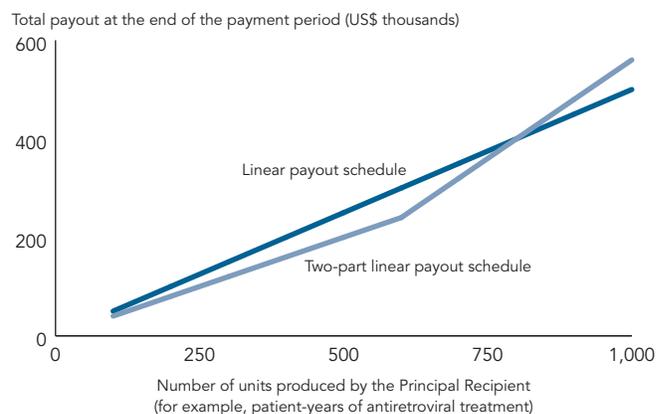
i. The baseline level could be defined as the level achieved on average during the three previous years. Or, in a situation where the level will change without substantial Principal Recipient effort, the baseline level could be defined by projecting the recent rate of change of the baseline forward to the end of the grant period—and then paying per unit of achievement beyond that baseline projection. For examples of payout schedules for HIV prevention, see Over (2011).

one or more targets and pay the Principal Recipient a lump-sum tranche for each target reached. However, experience at the World Bank and elsewhere suggests that such threshold payments can be vulnerable to serious complications. When measured performance falls just short of the threshold, country and donor officials can face strong pressure to fudge the verification process or payment decision to allow for disbursement. Coupled with Global Fund staff incentives that favor disbursement (described above), lump-sum payment schedules thus dilute the Global Fund's all-important power to commit. In most cases the Global Fund should mitigate these possible problems by using continuous payout schedules that reward the Principal Recipient for each incremental unit of progress (as long as doing so makes sense from a public health perspective). Continuous payout schedules lower the stakes of small differences in reported performance, helping the Global Fund resist political pressure to disburse within a binary decision framework (to disburse or do not disburse).

- *For continuous payout schedules—does the fixed price change with greater achievement?* The simplest payout schedule offers a single, constant price for each unit of output. For example, payment could equal the number of people living with HIV retained in antiretroviral therapy times a single price for each person-year of treatment—say \$500. If the Principal Recipient enrolled and kept 600 people on treatment, the payment would be \$300,000; if it added one additional person, payment would be \$300,500. This arrangement can be considered a linear payout schedule (figure 4.1).

But a simple linear payout schedule may not always be appropriate. For example, the Global Fund and Principal Recipient might agree that the cost and effort of service delivery will increase as the Principal Recipient reaches more patients—for example, expanding coverage beyond the low-hanging fruit in order to reach rural, marginalized, or otherwise costly clients. For example, a Principal Recipient may regularly enroll and retain 500–600 patients in antiretroviral therapy each year, at an average cost thought to be somewhere between \$300 and \$400 per patient. The Principal Recipient aspires to enroll and retain 800 or 1,000 patients in future years—but it recognizes that doing so will require substantial extra effort and extra outreach cost. The Global Fund could agree to pay \$400 per enrolled and retained patient up to the 600th patient and \$800 per patient

Figure 4.1. Linear and kinked payment schedules



thereafter—a kinked payout schedule (see figure 4.1). A third and more complicated option is described in box 4.3.

Setting and updating the fixed price

It can be very difficult to set the right price within a fixed price grant agreement.³² Too high a price will be wasteful; too low a price and the Principal Recipient will not have a clear performance incentive. And the criteria for the right price will vary based on several factors, including the purpose of the fixed price payment (supplement versus subsidy versus substitute for full cost, as discussed above), the setting, and policy goals. The challenge is particularly acute in settings where the fixed price is intended to substitute for input financing. There, the fixed price should closely mirror the actual cost of service delivery—but those costs may not be known at the time of grant negotiation and signature.

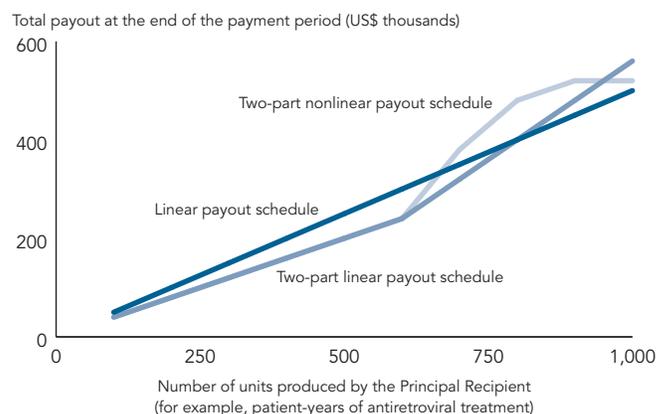
However, there are some practical approaches to determining the initial price and making adjustments over time as new information becomes available or in response to contextual changes. In particular, economists have proposed several ways to adjust prices over time as new information becomes available, input prices change, and the Principal Recipient is able to achieve gradual efficiency improvements.

First, the Global Fund could change the structure of the fixed price each year as it learns more about the costs incurred by the

Box 4.3 A nonlinear payout schedule to elicit information on efficiency

Another possible payment schedule would include a hill shape. It would give a low payment per unit to maintain baseline coverage, a larger payment for an initial expansion of coverage, and then declining incremental payments for subsequent patients, eventually falling below the price paid to maintain the baseline achievement (see figure). Such a payout schedule offers simplicity at low achievement levels, discretely higher incentives for exceeding the baseline level (for example, 600 units in the figure), and then a declining incentive beyond the target to elicit information on the efficiency improvements that the Principal Recipient can achieve with high scales of production. See Annex 2 of Glassman, Fan, and Over (2013) for a more detailed explanation of the benefits and the mechanics of this kinked nonlinear payout schedule.

A nonlinear payout schedule to elicit information on efficiency



Principal Recipient. In the first year it could reimburse the Principal Recipient for its actual costs. Then dividing by its output—for example, the number of people enrolled on antiretroviral therapy and retained for the first 12 months—the Global Fund could calculate a crude average cost. The next year, that average cost serves as a benchmark to set the fixed price, but the Global Fund would also track the Principal Recipient’s actual cost of providing that service—and that updated average cost would be paid as the fixed price in the subsequent year. Although the Principal Recipient might try to inflate costs in the first year, thereafter the Principal Recipient would be incentivized to exploit economies of scale and otherwise lower costs to achieve a surplus in each period, which it could retain and use for any health-related purpose. As the fixed price is adjusted from period to period, the Global Fund itself would also eventually benefit from that cost reduction. However, this mechanism would require the Global Fund to continue tracking at least a representative sample of receipts and expenses in order to observe the true cost of producing the service and regularly update its fixed price.

A second option would adjust the fixed price each year based on inflation (for example, changes in the costs of fuel, salaries, drugs, and so forth), less a predesignated efficiency reduction (see box 4.4 for a real world example). The Global Fund could also add an adjustment to account for changes in global commodities prices. This model would induce increased efficiencies or cofinancing by

the Principal Recipient over time while incentivizing increased output and protecting the Principal Recipient from global commodity price shocks.ⁱⁱ

Hedging your bets: Alternatives to simple fixed price mechanisms

While fixed price contracts are the easiest to understand, their downsides—risk to the Principal Recipient and the difficulty of setting the right price—may make them inappropriate in some settings. In those cases the Global Fund can still leverage some fixed price benefits through more sophisticated, modified mechanisms. These modified mechanisms insulate Principal Recipients from downside risk while still incentivizing them to achieve efficiencies and increase their productivity. There are two basic variations,

ii. A third option would exploit the fact that almost half of Global Fund grants are executed by Principal Recipients with more than five subrecipients (or subcontractors; see the lower right panel of figure 3.1). By collecting information on the cost and quality of comparable services delivered by each of several subrecipients, the Global Fund and the Principal Recipient could use the cost and quality of the best-performing subrecipient as a benchmark for the others.

Box 4.4 Dynamic fixed price models in the real world

Some dynamic models have already been implemented in real world settings. For example, following privatization of British Telecom in 1984, United Kingdom authorities created a new system to regulate the price of telecom services provided by the newly formed private sector monopoly. The regulators set a price ceiling for a core basket of telecom services; the ceiling was high enough for the company to make profits but low enough to protect consumers from exorbitant monopoly prices. Each year thereafter the maximum price for the basket was adjusted based on a predetermined formula. First, the price would increase by the retail price index in order to account for inflation. Second, the price would decrease by a set percentage rate—starting at 3 percent year and later rising as high as 7.5 percent a year—to induce gradual efficiency improvements and price reductions for British consumers.

Source: Green 1997; Pollitt 1999.

intended for slightly different circumstances: the fixed price/cost reimbursement menu and the shared surplus model.

Under the fixed price/cost reimbursement menu, the Principal Recipient is offered a choice between fixed price and cost reimbursement; the Principal Recipient does not have to decide between the two choices until the end of the payment period, when it knows which option will be more advantageous. This model is analogous to the choice of business travelers on their return between expense reimbursement or receipt of a flat-rate per diem (see box 4.5).

The fixed price/cost reimbursement model provides a good balance between the pros and cons of the fixed price and cost reimbursement models alone, particularly where the objective is to reach more people with a service and to do so more efficiently. If the Principal Recipient's costs turn out to be high, the Principal Recipient is free to seek cost-reimbursement up to a pre-agreed ceiling. However, the Principal Recipient has an incentive to try to keep costs low while increasing output, such that it can receive more fixed price payments and keep the surplus. Setting the right price is still challenging. If the fixed payment is too low, no Principal Recipient would ever choose it, leading to full cost reimbursement and no cost savings. If the fixed price is set too high, the Principal Recipients would

Box 4.5 Real world example: Travel per diems

After traveling on behalf of their companies, many business travelers face a choice between two options offered by their employers: request a flat-rate per diem allowance, set at a level that covers reasonable travel expenses, or submit all eligible receipts and receive reimbursement for the actual costs incurred.

For both the company and business travelers, this system offers several advantages. If travelers were always reimbursed for the actual costs, business travelers would want to eat gourmet meals and stay at the nicest hotels—driving costs up for the company. However, under this system business travelers are incentivized to keep their travel costs low and request the per diem—allowing them to pocket the surplus between the per diem amount and their actual costs. But they are also protected if the per diem amount is set too low or unanticipated costs arise—for example, if the less expensive hotels are fully booked or if prices are higher because of a big event or conference in the destination city. Thus this option provides some of the same benefits of a pure per diem system while protecting business travelers in the case of legitimate but higher than expected costs.

Over time, the per diem cost reimbursement system can also help the employer fine-tune the per diem amount. For example, the employer could gradually tweak the per diem amount for each city until that rate is selected by no more than 80 percent of employee travelers, signaling that the per diem rate is neither too high nor too low.

always choose it, yielding large benefits to the Principal Recipient but potentially increasing the cost to the Global Fund, at least in the short term. However, the Global Fund could then use that information to adjust the price upward or downward in future periods, approaching a more appropriate price point.

A second related model is often described as shared savings or shared surplus—and its application to the U.S. Medicare program has saved hundreds of millions of dollars (box 4.6). Under this model, the Global Fund would project the expected cost of either reaching a prespecified number of people with a defined package of services or of improving their health against some specific metric (for example, viral suppression). If the Principal Recipient is able to achieve that

Box 4.6 Real world example: Medicare shared savings program

In the United States, the Centers for Medicare & Medicaid Services have established a shared savings program. Under this program, the centers contract with coordinated groups of health care providers—accountable care organizations—and uses historical and projected trends to estimate their expected cost for providing high-quality health care to registered patients. The centers then pay the providers on a fee-for-service basis.

If accountable care organizations can keep their costs significantly below the projected sums (that is, outside a normal range of fluctuation) and maintain minimum quality standards, they are eligible to receive a lump sum share of the savings. The remaining amount is retained by the centers, such that both providers and taxpayers benefit from efficiency gains.

Accountable care organizations that exceed the cost projection, however, must pay the centers a portion of the difference—that is, shared losses. In the pilot program's first two years, covering just 32 accountable care organizations and about 800,000 patients, the program resulted in \$385 million in savings for the U.S. government, all without hurting patients' experience.¹

Note

1. Nyweide and others 2015; see Centers for Medicare & Medicaid Services 2014.

goal at lower cost, it can then keep a portion or the entirety of the surplus—for example, the difference between its actual costs and the expected costs of achieving those health outcomes. To incentivize high-quality service delivery, the Global Fund could also condition receipt of the surplus on quality metrics or on inclusion of marginalized communities in its health service provision.

Protecting communities, safeguarding rights

Many stakeholders rightly worry that the introduction of financial incentives could undermine the rights and welfare of citizens, and particularly communities of key affected populations for HIV and

tuberculosis—groups that are already underserved by existing services; marginalized by the societies in which they live; and vulnerable to violence, discrimination, coercion, and other rights abuses.³³ Appropriately, the promotion and protection of human rights are among the five Global Fund strategic objectives. The Global Fund is obligated to ensure that it “does not support programs that infringe human rights,” either in conventional or next generation grants.³⁴

For next generation grants, putting in place the wrong incentives can lead to one of two unacceptable scenarios. First, a Principal Recipient could respond to financial incentives by forcing community members to get tested or take treatment against their will —“processes...completely at odds with the rights-based approach which is the underpinning of effective HIV programs according to [the Global Fund], WHO, [and] UNAIDS and its co-Sponsors.”³⁵ Second, a Principal Recipient could determine that members of marginalized communities are too difficult or expensive to serve and cherry pick easier cases instead—thus failing to provide effective outreach or recruit them into testing and treatment or even actively discriminating against them in service provision.

But communities of key affected populations are often not adequately served through conventional approaches. For example, current prevention services are typically often offered by nongovernmental organizations, but there is little evidence that those activities lead to real protection against HIV and tuberculosis.³⁶ When HIV tests come back positive, the link to care—typically offered by the government—may be nonexistent. Done right, next generation grants thus offer an opportunity to greatly improve service coverage and quality for key affected groups.

Indeed, with the mitigation measures listed below, results-based grant designs offer an opportunity to considerably improve health service access and outcomes for members of key population groups while protecting their rights and autonomy. This section discusses three strategies to ensure that communities are served and their rights protected: indicator selection and disaggregation; community monitoring, feedback, and verification; and selection of an appropriate Principal Recipient and subrecipient.

- *Indicator selection and disaggregation.* Where possible, the very choice of indicator and the payment schedule should be designed to ensure that communities of key affected populations and other underserved groups—for example women, rural populations, or the very poor—are served. One strategy for achieving this is to disaggregate reporting and payment by population type.

Some grants could pay for services only to a specific population group; others could offer a higher payment amount for serving specific groups, accounting for the increased cost of doing so and providing a direct incentive to serve them. However, to protect against potential coercion and to ensure that providers do not misreport beneficiaries as belonging to a specific group in order to reap a higher reward, this strategy must be supplemented by community-based monitoring, feedback, and verification (described below).

- *Community monitoring, feedback, and verification.* To protect communities, it is vital that the communities themselves be a part of the process for monitoring, providing feedback, and verifying that results were indeed achieved—and that they were achieved in a way that was sensitive to community needs and did not involve coercion. Such a strategy could be multifaceted. First, there should be ongoing community monitoring and feedback mechanisms to report coercion or rights abuses, and those mechanisms should be accessible and anonymous. If human rights abuses are reported and confirmed, the grant should be immediately discontinued and replaced with input-based financing

to a different service provider to ensure continuity of services. Second, community-based organizations could be enlisted to help with community-based verification, sampling members of the community to ask whether they have proceeded through the full steps of the cascade; this can be cross-checked against provider records to validate the reported coverage.³⁷

- *Principal Recipient and subrecipient choice.* Where service delivery to key populations is a primary goal, a community-based membership organization should by default be given responsibility for recruiting new members; for offering prevention, testing, and counseling services; for acting as expert patients; and for measuring the size of the key affected community (the denominator needed to calculate coverage of service delivery). Having a key population membership-based group take on these roles is itself a protective measure, since these groups are directly accountable to the populations they serve.³⁸ Where the Principal Recipient is the government, the Global Fund should insert a clause in the grant agreement requiring the government to contract with one or more such groups as subrecipients.³⁹

Chapter 5

How you know: Verification, assurance, and assessment

*More Health for the Money: Putting Incentives to Work for the Global Fund and Its Partners*⁴⁰ extolled the importance of robust independent verification and identified four minimum requirements for its implementation: a technically sound and robust approach, unannounced visits, coverage of the core indicators tied to reimbursement, and the use of an independent third party to carry out verification activities. This chapter revisits these considerations and provide practical guidance on the design and implementation of a verification strategy.

In addition, it considers the broader question: how do you know whether the new grant designs are working more broadly? It thus proposes a simple assessment framework to guide assessment and facilitate learning.

Why verify?

Verification is a prerequisite for next generation grants. Under these models the Global Fund bases at least part of its payments to the Principal Recipient on the achievement of a specific set of outcomes, such as putting a certain number of patients on treatment. By itself, this creates a strong incentive for the Principal Recipient to misreport performance in order to increase disbursements from the Global Fund. This mismatched incentive cannot be addressed through assessments or improvements to the quality of Principal Recipient data systems alone, since even sophisticated Principal Recipients will be tempted to overstate their actual performance. Indeed, misreporting persists even in environments with highly sophisticated data systems—for example, overbilling in the U.S. hospital system.⁴¹ To align the Global Fund’s interest in accurate data with the Principal Recipient’s own interests in securing higher disbursements, it is thus incumbent on the Global Fund to verify the Principal Recipient’s self-reported performance and to apply penalties if misreporting or gaming is identified. Verification of self-reported performance is thus necessary both for accountability and fiduciary assurance.ⁱ

i. The Global Fund defines assurance as “The objective and independent review of grant activities by internal and external assurance providers with

The concern about potential misreporting and the resulting need for verification is not unique to the Global Fund. It arises in principal-agent relationships whenever it is difficult or costly to observe the actions of an agent that has incentives to misreport. There are many examples of agents misrepresenting their true performance and funders using verification to counter such behavior. For example, Sandefur and Glassman (2013) document overreporting of diphtheria-tetanus-pertussis vaccination coverage in administrative relative to independent survey data across several African countries. This may be driven by an incentive program from Gavi, the Vaccine Alliance; measles vaccinations, for example, were not incentivized and also not associated with systematic misreporting. Similarly, the U.S. government regularly examines whether claims submitted by hospitals are correct or inflated to assure that payments reflect actual and necessary medical care services.⁴²

Tradeoffs and credible threats

The principal-agent framework introduced above provides a useful way to characterize the incentives of Principal Recipients to misreport and potential mitigation strategies for the Global Fund. The fundamental problem is asymmetric information: the Global Fund is unable to costlessly observe the Principal Recipient’s true performance, which affords the Principal Recipient scope for misreporting. The Global Fund can use verification to establish the veracity of the Principal Recipient’s reported performance, but doing so is costly to the Global Fund.

The key practical issue is the tradeoff between the costs and precision of verification: verification efforts that are more involved or larger in scope are expensive but generally lead to more precise assessments of the Principal Recipient’s true performance. To resolve this tradeoff, the Global Fund can use verification policies

the ultimate goal of ensuring the achievement of each grant’s objectives.” Implementation of assurance reform was among the Global Fund’s eight corporate priorities for 2015. See Global Fund (2014a, 2014b).

that introduce a threat of detection and impose penalties for misreporting if detected. That is, the Global Fund can make clear at the outset that it will verify at least some of the Principal Recipient's self-reports and that evidence of misreporting will be costly to the Principal Recipient. Those penalties can take several forms, such as terminating a grant agreement, precluding access to future funding opportunities, publicizing the findings of verification, and imposing monetary fines. Alternatively, the Global Fund could withhold a small amount from the overall allocation and pay it out as a supplemental payment for accurate reporting.

Importantly, the findings of the verification exercise offer two distinct pieces of information that are relevant to the disbursement decision. First, what is the Principal Recipient's true performance that should be the basis of payment, as specified in the grant agreement? The true performance might be lower than the reported performance, thus implying a lower payment—but this is not a penalty, simply an accurate reflection of performance. Second, should the Global Fund apply a misreporting penalty—and, if so, how large should it be?

Resolving the tradeoff between costs and precision reveals several general important insights. First, the main goal of verification is not to fully eliminate misreporting, because doing so would be prohibitively expensive. Rather, verification aims to deter misreporting to within some tolerable range, where a smaller range (that is, higher precision) is associated with a higher cost of verification, requiring the Global Fund to balance the costs and benefits of greater precision during the design stage. Second, the Global Fund must impose a credible threat to verify the Principal Recipient's self-reported performance and to impose sanctions or penalties when misreporting is detected. Third, deterrence operates through the combination of the penalty level and the probability of detection. The Global Fund can contend with a lower probability of detection if the penalties are sufficiently high—so long as the expected value of cheating is lower than the expected value of playing by the rules. Fourth, verification should be implemented by an agency that is independent of the Principal Recipient and, to maximize the credibility of the exercise, probably also independent of the Global Fund.

Among these considerations, credibility stands out as critically important and costless to the Global Fund. Credibility matters in a single-country application, since the Principal Recipient would reduce misreporting only if it believes that verification will occur, that the verification will detect misreporting, and that sanctions will

be applied if misreporting is found. It may be even more important in the context of the Global Fund's work with multiple Principal Recipients, since lenience in one case may signal an institutional lack of credibility and weaken other Principal Recipients' incentives to report truthfully. Thus it is absolutely critical that the Global Fund and the Principal Recipient adhere to the ex ante contract terms without allowing ex post renegotiation.

Verification in practice

Probabilistic performance verification by an independent third party is critical both for deterring misreporting in the first place and for detecting misreporting after it has occurred. In any given situation, however, the specific design and implementation of a verification strategy will depend on the tradeoff between precision and costs (see box 5.1 for an overview of verification and evaluation in the Salud Mesoamerica Initiative). The overall question mirrors a well studied problem in survey sampling: how to achieve some desired precision at the lowest possible cost.

Once the Global Fund and Principal Recipient have agreed on contractible—and verifiable—indicators (see Chapter 3), there are several key parameters for designing a cost-effective verification strategy. This section discusses the following essential parameters: the instruments of verification; the frequency, timing, and scope of verification; and the consequences of discrepancies when detected.

Instruments of verification

Performance verification is distinct from evaluating the strength of a data system in that performance verification evaluates the accuracy of the information rather than the quality of reporting systems. (However, it is complementary in that it suggests whether the data systems are indeed recording the correct data values.) Specifically, verification instruments need to focus on capturing the true values for contracted outputs or outcomes. While common instruments such as the Data Quality Audit can be helpful tools for countries looking to strengthen their own national data systems, they are of limited use verifying performance because they do not mitigate Principal Recipient's fundamental incentive to misreport.

As noted above, there are several approaches to data collection that can be combined to balance different precision and costs. The mix of instruments to capture performance depends on the specific

Box 5.1 Verification in the Salud Mesoamerica Initiative

The Salud Mesoamerica Initiative's strong monitoring and evaluation component serves three main objectives: to provide reliable and independent data to verify results for the results-based financing model, to generate comparable and robust information for pro-poor policy dialogue, and to evaluate the initiative's model and key interventions.

The initiative created a Regional Results Framework to capture reproductive, maternal, and child health indicators using a lifecycle approach. Its indicators span the results chain, including impact measures such as the prevalence of anemia. Many of those indicators could be measured only using population-based approaches; the initiative thus conducted household- and facility-level surveys to measure and verify programmatic results.

Survey data were used to verify achievement of targets, but the data also proved essential for priority setting, program design, and policy dialogue. The baseline data generated many surprises. In many cases the poor were worse off than previously known, particularly with respect to quality of care, anemia, and stunting. Since the surveys were comparable across countries and externally verified, their data generated reputational incentives for countries to improve services for the poor.

context and indicators. For example, to verify indicators that rely on clinical data, such as HIV patients retained on antiretroviral therapy and virally suppressed, the Global Fund may need to conduct facility audits and chart reviews, supplemented by spot checks to trace patients in the community. Similarly, indicators that rely on population-level measures, such as access to treatment, may require household surveys among the target populations. Wherever possible, the Global Fund should seek to piggyback on existing or harmonized data collection and verification efforts rather than setting up new parallel systems; however, the Global Fund will need to ensure those existing tools offer sufficient rigor and independence to serve that purpose.

Although some modes of verification may require in-person interviews or visits, the proliferation of mobile phones may allow lower cost data collection, especially among hard-to-reach or

scattered target groups. Although phone or SMS surveys may require shorter questionnaires, they can be used to reach more respondents, possibly rotating questions so that, overall, all questions can be asked of the target group. Phone surveys have already been deployed in several settings; in Tanzania researchers found that 2,677 phone survey interviews cost under \$20,000 in 2010.⁴³

Frequency, timing, and scope of verification

The frequency, timing, and scope of verification are critical elements of the overall strategy because they are components—paired with the penalty—of the overall deterrent effect. Since the findings will be compared against reported performance, it is sensible to align the verification frequency and timing with the agreed schedule for Principal Recipient reporting and results-based disbursements. This could be done through small annual verification exercises, together with a larger assessment to coincide with the three-year grant cycle.

The smaller but regular verification visits should be random and unannounced to maintain an effective deterrent effect. That is, the Global Fund should not inform the Principal Recipient which facilities or communities will be visited when. Since the Principal Recipient cannot anticipate which facilities or communities will be checked, the deterrent effect applies across the Principal Recipient's entire portfolio. This principle can also be applied to scenarios with multiple performance indicators. In such circumstances the Global Fund may not need to verify every indicator but could randomly select a subset.

The larger and less frequent assessments could draw on existing efforts to collect population-level data, such as Demographic and Health Surveys, or facility surveys such as Service Provision Assessments. These surveys are implemented independently of the Principal Recipient's program and can therefore provide a robust benchmark against which to compare the Principal Recipient's reported performance. In addition, they impose no additional cost or burden on the Global Fund or Principal Recipient, since they are already conducted as part of routine business.

One important feature of verification in the context of the Global Fund is that the Global Fund should not care (for the purposes of payment) whether the Principal Recipient's actual performance exceeds its reported performance, since this means the Global Fund is actually underpaying for the results achieved. This has implications for designing the verification sample, which can be smaller

than if the Global Fund was interested in assessing both over- and underreporting.

So we found a discrepancy: Now what?

As noted above, verification findings should be used to inform two distinct components of the funding decision: appropriate payment for the true performance and a punitive penalty for the misreporting itself. The role of the former is to ensure that disbursements match actual performance. Given that the Principal Recipient did not perform some services—even though it had reported it had—there is no basis for payment. The function of the penalty, in contrast, is to disincentivize deliberate misreporting. In practice, the penalty for misreporting could also be the nonpayment of an additional bonus for accurate reporting. In both cases the Principal Recipient would have a financial incentive to report accurately, since failing to do so would entail a financial loss.

In a perfect world the Global Fund would be able to determine the right amount for each of the two payment components based on the true value of performance and the true extent of misreporting. But since absolute precision is generally neither feasible nor cost-effective, the grants between the Global Fund and the Principal Recipient need to account for uncertainty and to avoid unwarranted punishment of the Principal Recipient based on statistical noise. This involves *ex ante* agreement at grant signature on how imprecise verification findings will inform disbursements and penalties—and requires the Global Fund to follow through on those agreements to maintain its credibility.

What does this mean in practice? Once the self-reported value is in hand and the verification strategy is known, the Global Fund can implement the verification to calculate the point estimate of the verification statistic and an uncertainty interval around it. The uncertainty is a consequence of the probabilistic sampling. Subsequently, the Global Fund can compare the self-reported performance with the point estimate and uncertainty interval from the verification. This is akin to a statistical test with the null hypothesis that the Principal Recipient's self-reported value accurately reflects true performance.⁴⁴

At this point, as shown in figure 5.1, there are several possible approaches. Importantly, these decisions and the associated contingencies must be clearly specified, *ex ante*, so as to avoid *ex post* renegotiation of terms.

In the first scenario self-reported performance is below the verified amount—that is, the Principal Recipient underreported its performance. In that case the Global Fund should disburse according to the Principal Recipient's self-reported performance, incentivizing it to strengthen its data systems and report more fully in future cycles. In the second scenario the self-reported performance falls within the uncertainty range of the verification. The Global Fund cannot reject the null hypothesis that the Principal Recipient report is accurate—so it should also disburse according to the self-reported performance.

In the last scenario (3a and 3b) the Principal Recipient's self-reported performance exceeds the uncertainty range of the verification point estimate. In this scenario there are two possible approaches. The first approach (3a) is to use the verification point estimate as the best estimate of true performance; that would imply that disbursements and punitive sanctions should be based exclusively on the verification point estimate. The second approach (3b) is to give the Principal Recipient the benefit of the doubt within the range of uncertainty—that is, to base the disbursements and punitive sanctions on the upper bound of the uncertainty range around the verification point estimate. The right choice between these two is ultimately a value judgment—but the decision rule must be negotiated before grant signature and explicitly noted in the grant agreement.

Learning while doing: Assessment and evaluation

Moving from the current funding approach toward next generation models will involve risks but will also provide opportunities for learning. A structured assessment framework can help the Global Fund determine whether the new models are working with respect to its own objectives and whether and how to tweak the models to improve their effectiveness and impact. The assessment framework should look primarily at changes in process but—where possible—should also aim to directly examine the impact of the new grant structure on beneficiaries.

New grant designs should be assessed with respect to the stated goals for their operationalization and pre-stated hypotheses about how they should lead to change along four dimensions: the Global Fund's behavior, the Principal Recipient's behavior, the interactions between the Global Fund and the Principal Recipient, and the impact on the health and welfare of program beneficiaries.

Figure 5.1. Verification scenarios and associated disbursements to the Principal Recipient

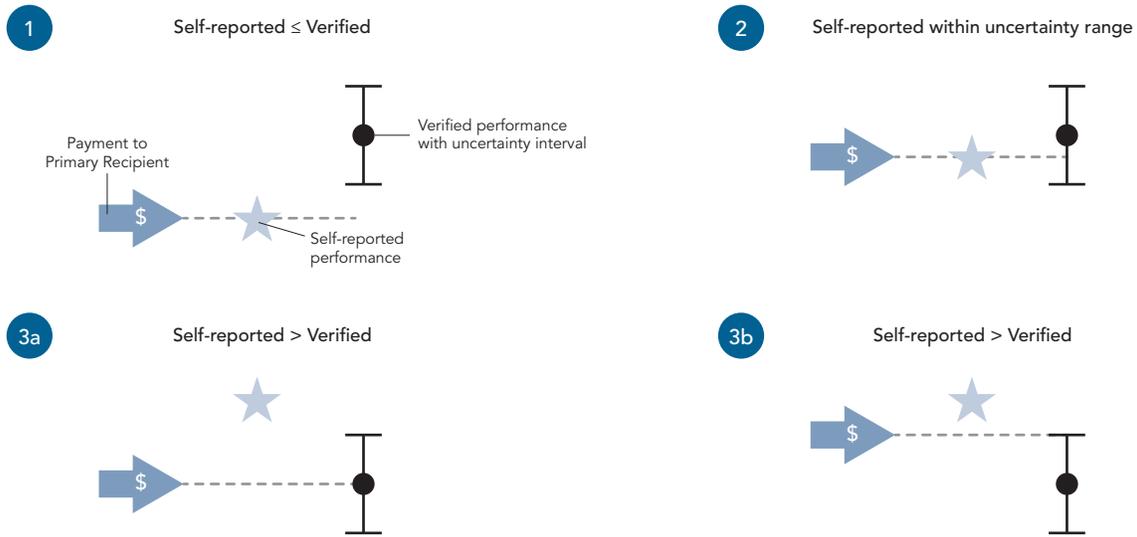


Figure 5.2. Sample assessment framework—Hypotheses and risks throughout the grant cycle

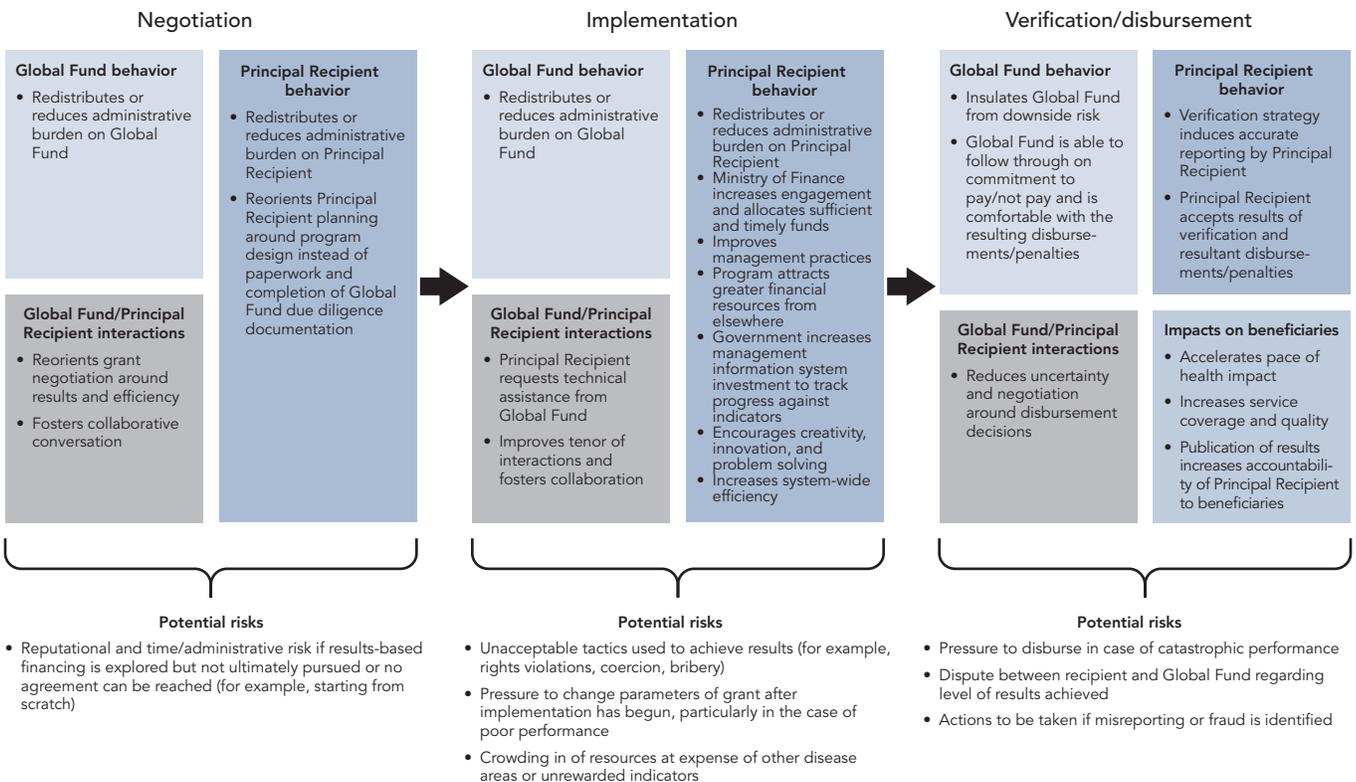


Figure 5.2 presents a sample assessment framework with hypotheses along each of those four dimensions at various stages in the grant cycle. For example, in some cases a grant might be said to work well if it leads to innovation in program delivery and implementation

or increases the efficiency and effectiveness of programs in achieving the designated outcomes—and if it does so without leading to unintended negative consequences above some predetermined threshold of acceptability.

Part III

Putting it all together

Chapter 6

Country case studies

This chapter applies the framework presented in Part II to four hypothetical country case studies: HIV in a generalized epidemic, tuberculosis, malaria, and HIV in a concentrated epidemic.

Each case draws from the real dynamics and challenges faced in those specific settings. The structure below reflects a rigorous analysis process to determine the appropriate model.

- Step 1: Understand the situation and context.
- Step 2: Define the policy objectives—and consider whether a different model could help achieve those policy objectives.
- Step 3: Review strategic considerations, risks, and constraints that may influence the choice of model or proportion of funding allocated to results payments, including:
 - Whether policy objectives can be measured using contractible indicators.
 - Whether contractible indicators are measurable and verifiable in that context.
 - Principal Recipient’s ability to prefinance activities.
 - Whether potential benefits of new model would outweigh costs and risks.
- Step 4: Define the proportion of funding allocated to results payments.
- Step 5: For the results-based payments, define the contractible indicators, payment mechanism, and verification strategy.
- Step 6: Explicitly define unresolved issues that need to be addressed during the negotiation period.
- Step 7: Review entire grant model and ensure that individual elements and the model in its entirety respond appropriately to the policy objectives, address contextual constraints, and mitigate possible risks.
- Step 8: Assess and evaluate the model.

Case 1: HIV in Country A

Situation analysis

Country A is a large, low-income country in Sub-Saharan Africa. Its health sector depends heavily on donor financing; total expenditure on health is about \$50 per person, including about \$10 in domestic public expenditure. Country A has a generalized HIV epidemic; 1 million citizens are living with HIV, and roughly a third of them are enrolled on antiretroviral therapy. Country A struggles with low rates of antiretroviral therapy retention (65 percent after one year) and major gaps in antiretroviral therapy. In 2013 an estimated 50,000 citizens of Country A died of AIDS.

For 2014–16 Country A received an HIV allocation of \$350 million from the Global Fund, which should be complemented by roughly \$1 billion from the U.S. President’s Emergency Plan for AIDS Relief and \$200 million from the Country A government.

The Global Fund’s contribution pays primarily for antiretroviral therapy; U.S. President’s Emergency Plan for AIDS Relief money is more widely distributed across both treatment and prevention.

Principal Recipient

The Country A Ministry of Health

Policy objective

To increase enrollment and retention in antiretroviral therapy for eligible people living with HIV.

Strategic considerations

Country A is a large country with a major HIV epidemic—and despite substantial donor investments, quality and coverage of antiretroviral therapy remain low. The Global Fund pays primarily for health commodities, giving it few levers to push for improved

antiretroviral therapy coverage or to influence the quality of service delivery. Because Country A is a low-income country, it has little capacity to prefinance its HIV response. And because the Global Fund contribution pays mostly for lifesaving antiretroviral drugs, the Global Fund faces ethical constraints in withholding its funds due to poor performance.

A next generation financing model should respond to this situation by incentivizing increased antiretroviral therapy enrollment and retention. But design elements must be responsive to the Country A context, where monitoring and evaluation systems are weak, most funds are tied up in commodity procurement, the Global Fund shares service delivery costs with the U.S. President's Emergency Plan for AIDS Relief, and the government is unlikely to offer much prefinancing.

Design elements

Proportion results based: 10 percent (\$35 million). The remainder of funding (including for all commodities and drugs) is delivered using traditional grant mechanisms.

Indicator: Number of people newly enrolled in antiretroviral therapy and retained for at least 12 months (using the World Health Organization definition of minimum number of visits as the basis for retention).

How to pay: In year 1 the Global Fund makes a fixed price incremental payment for each newly enrolled patient during the grant's first six months who has complete and accurate retention records at the end of the year, up to the 10 percent ceiling.

In year 2 the Global Fund makes a fixed price incremental payment for each person enrolled during year 1 who has been retained in treatment after 12 months, up to the 10 percent ceiling.

In year 3 the Global Fund makes a fixed price incremental payment for each person enrolled during year 2 who has been retained in treatment after 12 months, up to the 10 percent ceiling.

Verification: The Global Fund will conduct unannounced verification visits in advance of disbursement decisions, ideally through an existing facility survey such as the U.S. President's Emergency Plan for AIDS Relief's Site Improvement Monitoring System. Visits will

be to a random sample of facilities; at each facility, verification activities will include a chart review for a random sample of patients. In some facilities it will triangulate charts with in-person spot checks to verify chart accuracy. In addition, it will incorporate a hotline where patients can report abuse, stock-outs, or service interruptions—a social accountability mechanism.

Remaining areas for negotiation:

- Size of the fixed price payment for each new patient enrolled and retained on antiretroviral therapy.
- Whether to set differential fixed price payments to incentivize enrollment and retention in specific groups or geographic areas.
- Sample size for verification: number of facilities, charts, and patient spot checks.
- Choice of verification agent—can the Global Fund piggyback off of the U.S. President's Emergency Plan for AIDS Relief's effort?
- Size and application of the penalty for overreporting.
- Restrictions or guidelines for fixed price payment spending.

Rationale for grant design

The proposed indicator is designed to incentivize progress against a major policy objective: faster enrollment and better retention of individuals newly living with HIV in antiretroviral therapy treatment. The relatively small amount of the payment is responsive to the constrained environment, where the low-income Country A government cannot cover the cost of prefinancing, and where most Global Fund monies are spent on antiretroviral drugs that cannot be discontinued in the case of poor performance. Thus the conservative design works to strengthen the current system rather than to overhaul it. As a supplemental payment, it is not intended to cover the full cost of service delivery—but it does push Country A to achieve quality gains in the programs that are already being funded from other sources.

The count indicator has several advantages in this case. First, it incentivizes incremental progress—each new person enrolled can generate a bigger payment. Second, it incentivizes improved recordkeeping—the government can claim the supplemental payment only if there is a complete and accurate record that a new patient was enrolled and retained on treatment. Third, it incentivizes what matters—not just enrollment, but enrollment paired with high-quality service delivery that keeps patients in care.

Case 2: Tuberculosis in Country B

Situation analysis

Country B is a small, lower middle-income country in Central Asia with a population of 10 million and GDP per capita of \$2,000. Total health expenditure is \$100 per capita—two-thirds of which comes from government coffers. Tuberculosis is a major health problem. Country B's tuberculosis incidence rate is estimated at 200 cases per 100,000 people—one of the highest rates in the region. The case detection rate is estimated at just 60 percent. A high proportion of tuberculosis cases—35 percent among new smear-positive patients—are multidrug resistant. The high rate of multidrug resistance reflects frequent treatment interruption and insufficient measures to support patient adherence. Just 70 percent of new smear-positive tuberculosis patients successfully complete treatment—well below the WHO target of 85 percent.

A major driver of the epidemic is the nosocomialⁱ spread of tuberculosis, including transmission of drug-resistant strains. Country B's hospital payment system ties remuneration to the number of days spent in hospital beds, incentivizing lengthy and unnecessary hospitalizations even when a patient could be effectively managed in a community setting. Integrating tuberculosis treatment into primary healthcare remains a major challenge.

For 2014–16 Country B received a tuberculosis allocation of \$20 million from the Global Fund. The Global Fund contribution is supplemented by an anticipated government allocation of \$60 million over the same period. Historically, a majority of the Global Fund contribution has supported tuberculosis treatment. Some 75 percent of Global Fund costs are associated with commodity purchases and supply-chain management.

Principal Recipient

Country B National Tuberculosis Program

Policy objectives

- To increase case finding.
- To improve treatment success rates.
- To reduce the proportion of cases treated in hospital settings.

Strategic considerations

In this setting the historical allocation of Global Fund financing does not directly align with its strategic interests. Funding is concentrated on commodities for tuberculosis treatment, particularly first- and second-line drugs. This gives the Global Fund few levers to push forward its primary goal: preventing the further spread of drug-resistant strains by improving first-line treatment success rates; improving case finding; and reducing the proportion of cases treated in hospital settings, which drives nosocomial tuberculosis transmission. In addition, the Global Fund acts as a minority funder; its contribution is supplemented 3:1 by government funding, often to counterproductive ends (for example, subsidizing unnecessary hospital stays). Thus the challenge for the Global Fund is to use its contribution to leverage changes in the health financing system writ large and to increase effort from the health sector toward improved case detection and better adherence support.

Design elements

Proportion results based: All noncommodity funds—25 percent of the total grant amount. Commodities would continue to be purchased directly by the Global Fund.

Indicators:

- Number of bacteriologically positive pulmonary tuberculosis cases detected in an ambulatory setting.
- Number of bacteriologically positive pulmonary tuberculosis cases that successfully complete first-line treatment in an ambulatory setting and achieve culture conversion.
- Proportion of bacteriologically positive pulmonary tuberculosis cases treated in an ambulatory setting.

How to pay: Results-based payments would be made in three annual installments, following each year of the three-year grant cycle. A total of \$5 million would be available over the course of the grant cycle (for example, 25 percent of the total allocation amount).

First, the Principal Recipient would receive a fixed price payment for each bacteriologically positive pulmonary tuberculosis case detected in an ambulatory setting, up to a maximum of \$1 million. The fixed price would be determined in negotiation with the Principal Recipient.

i. That is, originating in a hospital.

Second, the Principal Recipient would receive a fixed price payment for each bacteriologically positive pulmonary tuberculosis case that successfully completes first-line treatment in an ambulatory setting and achieves culture conversion, up to a ceiling of \$3 million. In the first year the payment would cover the cohort of patients enrolled during the first six months of the grant. In the second year the payment would cover the cohort of patients enrolling between months 7 and 18. The final payment would cover the cohort of patients enrolling between months 19 and 30.

The final component would be cash on delivery payments of up to \$1 million for increasing the proportion of tuberculosis cases managed in an ambulatory setting. The baseline would be measured and verified at the beginning of the grant cycle. Thereafter, the Principal Recipient could access payments on a continuous scale for percentage point increases at \$40,000 per percentage point increase, up to the full \$1 million for a 25 percentage point improvement.

Verification: A verification agent would visit a stratified random sample of facilities (stratification would be based on caseload) and conduct chart review of patient records. Patient charts would then be triangulated against laboratory and drug dispensary records and supplemented by spot checks to verify the existence of tuberculosis patients and accuracy of triangulated records.

Remaining areas for negotiation:

- Size of the fixed price payments for case finding and treatment completion.
- Sample size for verification: number of facilities, charts, and patient spot checks.
- Choice of verification agent.
- Size and application of the penalty for overreporting.
- Restrictions or guidelines for fixed price payment spending.

Rationale for grant design

The Global Fund considers itself ethically obliged to ensure a continuous supply of treatment commodities, thus precluding it from tying commodity contributions to the Principal Recipient's performance. After accounting for those commodities, the Global Fund would be left with about a quarter of its total allocation—about \$5 million over three years—that it could use to leverage improvements in its priority areas. And because the Global Fund's noncommodity contribution represents just a small portion of Country

B's overall financing for tuberculosis, it is reasonable to expect the Principal Recipient to prefinance tuberculosis service delivery with the expectation of later Global Fund payments. Using a modular approach, this next generation financing model thus addresses three discrete Global Fund objectives: improved case finding, treatment completion, and community-based care.

Each indicator serves multiple objectives, and together the three indicators are mutually reinforcing. The first part of the payment encourages providers to find more tuberculosis cases, even if doing so requires additional effort or expense. By requiring a laboratory confirmation, it also incentivizes stronger laboratory capacity and broader coverage of high-quality diagnostic services. And because it applies only to bacteriologically-positive cases, it incentivizes providers to find the sickest patients and discourages misdiagnosis on the basis of clinical exam alone.

The second indicator likewise incentivizes multiple positive outcomes. It reinforces the incentive for better case finding, since each case later becomes a candidate for a treatment success payment. It encourages providers to support patient adherence during treatment, since the providers will receive incentive payments only for patients who complete treatment. Finally, the requirement that treatment be delivered in community settings is itself a partial disincentive to hospital-based care, perhaps at least in part counterbalancing the perverse incentive introduced by Country B's bed-based hospital financing system.

Finally, the third indicator—the proportion of cases treated in an ambulatory setting—is intended as a small but visible push for the government to achieve a long-overdue policy change. The policy change would cost the government nothing—indeed, it would be cost saving—but would require political will and resolve to address effectively. The payment would be designed to increase political engagement around this policy question, hopefully prompting faster progress than through the status quo.

Case 3: Malaria in Country C

Situation analysis

Country C is a lower middle-income country in Sub-Saharan Africa with a population of 25 million. Total health spending there is roughly \$100 per capita, including about \$50 per capita from external sources. Malaria has historically been among the leading causes of morbidity and mortality in Country C, particularly among young

children and pregnant women. In the last decade, with considerable support from the U.S. President's Malaria Initiative and the Global Fund, Country C has made great strides in scaling up vector control and increasing access to artemisinin-based combination therapy, leading to substantial declines in all-cause child mortality.

Despite this progress, Country C has struggled to scale up intermittent preventive treatment of malaria for pregnant women. Although a large majority of pregnant women (85 percent) attend at least four antenatal visits, just 40 percent receive the recommended three doses of intermittent preventive treatment of malaria for pregnant women. The largest bottleneck to further scale-up is health worker confusion about guidelines for administration.

For 2014–16 Country C received a malaria allocation of \$80 million from the Global Fund. The funds are supplemented by \$20 million a year in U.S. President's Malaria Initiative support, \$8 million a year from the United Kingdom's Department for International Development, and \$25 million a year in domestic funding for the National Malaria Control Program.

Principal Recipients

- National Malaria Control Program of Country C.
- Ministry of Health of Country C.

Policy objective

To increase scale-up of intermittent preventive treatment of malaria for pregnant women among pregnant women at risk.

Strategic considerations

Under the current financing model, most elements of the Global Fund grant are working well. The National Malaria Control Program has proven effective and efficient in scaling vector control interventions, and Country C has adopted and rolled out artemisinin-based combination therapy as its first-line malaria treatment. There is no reason to change the basic financing arrangement for the majority of the grant.

However, intermittent preventive treatment of malaria for pregnant women stands out as a stubborn challenge within an otherwise successful program. Intermittent preventive treatment of malaria for pregnant women is delivered by the public health sector during routine antenatal care, under the auspices of the Ministry of Health but outside the direct control of the National Malaria Control Program. Increasing coverage of intermittent preventive treatment of malaria

for pregnant women is a management and training challenge; the drugs are already available, but many health workers remain confused about the guidelines for their administration, in part due to a relatively recent change in World Health Organization guidance.

To improve coverage, the Global Fund must mobilize the Ministry of Health to train health workers on the appropriate administration and take measures to ensure compliance with the guidelines. However, studies from elsewhere in Africa have shown that expanding intermittent preventive treatment of malaria for pregnant women can at times lead to decreases in coverage of antenatal care.⁴⁵ The Global Fund must ensure that scale-up of intermittent preventive treatment of malaria for pregnant women does not unintentionally contribute to decreases in coverage for other elements of antenatal care.

Design elements

Proportion results based: 5 percent of total grant value (\$4 million).

Indicator: Percentage of women at risk who receive three or more doses of intermittent preventive treatment of malaria for pregnant women and attending at least four antenatal visits during their last pregnancy, calculated at end of three-year grant cycle by a previously scheduled Demographic and Health Survey.

How to pay: Following administration of the Malaria Indicator Survey in year 3 of the grant, the Global Fund would pay the Ministry of Health a flat sum of \$100,000 per percentage point increase in coverage over baseline, up to \$4 million for a 40 percentage point increase.

Verification: Calculated by representative household survey.

Remaining areas for negotiation: None.

Rationale for grant design

This design targets the one stuck component within an otherwise successful conventional grant. The results-based portion is supplemental to input financing and represents just a small portion (5 percent) of the overall value. To mitigate the risk that some women might substitute standalone intermittent preventive treatment of malaria for a full package of antenatal care, the indicator requires the women to receive both intermittent preventive treatment of malaria and to attend at least four antenatal care visits.

Given the small size of the supplemental incentive, it is desirable for the Global Fund to avoid complexity, costly verification requirements, or other high transaction costs. This design is appropriate given that context. First, it piggybacks onto an existing measurement exercise, requiring no additional monitoring and evaluation or verification investment. Second, the payout function is clear and simple, rewarding incremental coverage improvements over a pre-established baseline. Finally, the incentive targets the Ministry of Health, even though the National Malaria Control Program is the Principal Recipient for the remainder of the grant. This feature is intended to mobilize the parties responsible for service delivery of intermittent preventive treatment of malaria for pregnant women, which occurs during routine antenatal care visits administered by public health facilities.

Case 4: HIV in Country D

Situation analysis

Country D is a midsize country in Latin America with a population of about 20 million, GDP per capita of \$5,000, and total per capita health expenditure of \$500.⁴⁶ Country D's health sector is funded primarily by domestic revenue; in 2013 public health expenditure accounted for almost 75 percent of total health spending, and external resources for health accounted for just 2 percent of total expenditures on health. Country D has a concentrated HIV epidemic; although prevalence in the general population is well under 1 percent, prevalence is far higher among men who have sex with men (20 percent). About half of new HIV infections occur among men who have sex with men. Just 1 in 10 men who have sex with men received an HIV test in the last year and knew their results.

Country D has an equitable legal regime, and men who have sex with men are not criminalized. Nonetheless, Country D has not been sufficiently proactive in reaching men who have sex with men with needed health services. Country D does not currently and has no plans to implement a test and treat policy for key populations.

For 2014–16 Country D received an HIV allocation of \$8 million from the Global Fund. These funds complement roughly \$50 million a year in government funding for the HIV/AIDS program.

Principal Recipient

Country D Ministry of Health

Policy objective

To scale up high-quality, rights-based, and sustainable HIV testing and treatment services for Country D's community of men who have sex with men.

Strategic considerations

Country D is an upper middle-income country with a relatively small HIV epidemic—and the Global Fund is likely looking to transition Country D away from its support following this three-year funding cycle. As a result, sustainability—and particularly sustainable, rights-based service delivery to key population groups such as men who have sex with men—is of paramount importance. Here, rights-based specifically refers to the right to health and to ensure that men who have sex with men have easy and fast access to necessary health services, including antiretroviral therapy to achieve viral load suppression. Finally, given Country D's relative wealth, it is reasonable to expect that the government could prefinance service delivery of Global-Fund-supported interventions.

Design elements

Proportion results based: All except small monitoring and evaluation investment costs to establish a patient tracking system and ombudsman hotline.

Indicators:

- Number of men who have sex with men tested for HIV and received the result (calculated biannually).
- Number of men who have sex with men living with HIV enrolled on antiretroviral therapy (calculated biannually).
- Number of men who have sex with men living with HIV achieving viral load suppression (calculated annually).
- Country D expenditure on HIV prevention services for men who have sex with men remains constant or increases.ⁱⁱ

How to pay: The Ministry of Finance would be required to forward funds to the Ministry of Health for the prefinancing of service delivery; the Ministry of Health would then be required through

ii. This indicator assures that Country D does not simply divert prevention funds toward treatment or otherwise deprioritize prevention efforts due to the new incentives for treatment outcomes.

its agreement with the Global Fund to forward a portion of those funds to contract nongovernmental organization service providers for outreach. This would include sufficient financial resources for clinical and diagnostic services to treat up to 36,000 men who have sex with men living with HIV. The Global Fund would then provide payments to the Ministry of Health based on each of the three indicators described above. Each indicator would have a different fixed price; the price would be negotiated between the Ministry of Health and the Global Fund taking into account the following considerations:

- Current costs of outreach to men who have sex with men, scaled up to reach 36,000 men who have sex with men living with HIV.
- Current costs of clinical and diagnostic services.
- Costs of mechanisms to ensure collaboration between government clinics and nongovernmental organization or network of men who have sex with men outreach staff.
- Costs to address barriers to services including stigma, discrimination, and other human rights issues.

Verification: The Global Fund will assist Country D in designing a patient tracking system for voluntarily self-identified men who have sex with men using anonymous unique identifiers. Verification will focus on whether the men reported as men who have sex with men are really voluntarily self-identified men who have sex with men (as opposed to men who do not have sex with men). Verification will be conducted by an independent consultant or institution with the support of a national or regional network of men who have sex with men in a stratified random sample of facilities; no prior notice will be given. The verification agent will check records of antiretroviral therapy medications received by the clinic against reported doses distributed to patients and current stock. The verification agent will also select (using a random number generator) a random sample (2 percent) of men who have sex with men clients; the agent will meet with them one on one to determine whether their patient records accurately reflect their number of visits, doses, and so forth, and to assess the quality of services provided. In addition, the Global Fund will support the national network of men who have sex with men to set up and publicize a peer-led hotline, enabling men who have sex with men to report complaints, coercion, or manipulation—for example, if they receive a request to re-enroll under a false or different name. This effort would be supplemented by the involvement of peer mediators, who could help connect men who have sex with men with the hotline or relay their concerns. Any reports to the hotline will be quickly and thoroughly investigated.

Remaining areas for negotiation:

- Size of the fixed price payments for testing, enrollment, retention, and viral load suppression.
- Sample size for verification: number of facilities, charts, and patient spot checks.
- Choice of verification agent.
- Size and application of the penalty for overreporting.
- Consequences of complaints to the hotline.

Rationale for grant design

This grant builds from a rights-based foundation to address the HIV outreach and treatment needs of Country D's men who have sex with men community—an essential step in curbing Country D's HIV epidemic. The service provision covered under the grant—and the incentives created by the three selected indicators—pair well with the largest drivers of the national epidemic: incomplete testing coverage and poor antiretroviral therapy retention among men who have sex with men. The incentives cover three essential steps in the cascade: testing, link to care, and viral suppression through retention in care. And the incentives build on each other; the government can only access Global Fund viral suppression incentives if it first works closely with the men who have sex with men community, tests men who have sex with men with informed consent, informs them of their status, and provides them with antiretroviral therapy for sufficient time and at appropriate doses to achieve viral load suppression. The final indicator is binary, ensuring that the country maintains or increases its current level of prevention funding; this prevents the country from diverting scarce prevention resources toward the incentivized treatment outcomes.

In addition, given Country D's upper middle-income economy and relatively small HIV epidemic (relative to its population), the grant is designed to promote sustainability and prepare the country for transition away from Global Fund financing—all while ensuring the continuation of services to key, marginalized populations. In requiring the Ministry of Finance to prefinance service delivery, the Global Fund will get its programs on-budget, a crucial step for sustainability. And in creating a structure where the Ministry of Health subcontracts to the nongovernmental organization service providers—and where it has a vested interest in their success—the grant promotes and strengthens government and civil society cooperation to deliver essential community-based health services.

Part IV

The way forward

Chapter 7

Recommendations for the way forward

This chapter offers key strategic and operational recommendations for the Global Fund Board and Secretariat to move forward with next generation financing models in the context of its forthcoming strategic framework and replenishment. These recommendations have been discussed with the Global Fund; however, they represent the opinions of the authors and CGD co-chair alone.

1. Secure strong Board and Secretariat commitment through inclusion of next generation grants as a key priority within the next Global Fund Strategy (due to be presented to the Board for approval in April 2016).

As this report has previously cautioned, the current evidence base on results-based grant designs is limited. Nonetheless, given the weight of the theory and evidence, we believe there is sufficient reason to believe that next generation grants—if wisely designed and implemented—can increase the pace of progress against the Global Fund’s three target diseases compared to its conventional approach. We thus recommend that the Board and Secretariat commit to the gradual rollout of next generation grant designs, coupled with careful evaluation to assess their impact and inform future expansion. That commitment should be reflected in the forthcoming the next Global Fund Strategy, endorsed by the Board, and publicized in advance of the Global Fund’s 2016 resource replenishment.

2. Leave no room for ambiguity: ensure that next generation grant agreements stick to their agreed disbursement protocols—against progress on independently verified results.

Part III laid out the various design elements that make up next generation grant agreements. Here, we emphasize that the grant

agreement itself must be the explicit and final word on all design elements. Leaving details ambiguous or allowing for midterm changes is problematic, since it dilutes the incentives of both parties. Thus the grant agreement must specify—with sufficient detail—the indicators; the payout schedule, including for catastrophic contingencies; the verification arrangements; the consequences of verification findings; the level of grantee flexibility over financial, procurement, and managerial decisions; required financial and other documentation; and any other elements of the design.

Once those design decisions are made, but before the next generation grant agreement is signed, it is essential that both the Global Fund and Principal Recipient understand—and accept—how all different possible scenarios will affect disbursements. The Global Fund should explicitly map out the consequences of its proposed design under different plausible scenarios, ensuring that all possible outcomes are understood and accepted by both parties. It may also consider offering technical assistance to help the Principal Recipient understand and negotiate the proposed grant agreement. If a specific scenario is unacceptable to either party, the grant design should be tweaked before being signed.

Once the grant is signed, the Global Fund must commit to using the grant parameters as the final guidelines for calculating the payment—and disbursement decisions must be insulated from political pressure. To do so, we recommend that the Global Fund separate its disbursement function from grant design. For example, the Global Fund could create an independent team to execute disbursement decisions for results-based payments on the basis of the pre-agreed grant parameters. Doing so would protect the Country Teams from conflicts of interest (for example, pressure from their in-country counterparts to disburse) while incentivizing both the Principal Recipient and Country Team to sign a grant agreement that offers complete ex ante clarity.

3. Reflect the needs and requirements of next generation grants in relevant related policies, including the allocation formula, counterpart financing requirements, sustainability framework, and differentiation initiative.

Adoption of next generation grants will have implications for several related Global Fund policies—specifically, the allocation formula, counterpart financing requirements, sustainability framework, and differentiation initiatives. All such policies should thus reflect the needs and requirements of next generation grants, helping create a cohesive policy framework across the entirety of the Global Fund.

At present, each country receives a set allocation from the Global Fund, which it can access once within the three-year funding window. Because this amount is effectively set aside for the country—and because the Global Fund expresses its intent to ensure grant implementation success—countries may perceive those allocations as entitlements and see any withholding of funds below that ceiling as a punishment. This structure presents a challenge for results-based payment designs, since it makes withholding payments politically unpalatable.

Wider adoption of next generation financing mechanisms may thus require the Global Fund to rethink both its allocation formula and its framing of country allocations. For example, the Global Fund could consider putting aside 10–20 percent of its overall budget for supplemental funds—that is, funds that could be accessed as an additional payment on top of country allocations in exchange for exemplary performance. Alternatively, it could offer countries a higher allocation in exchange for accepting the increased risk of a new financing modality. It could also explicitly communicate that allocations are not entitlements—particularly in transitioning countries—and note that receipt of those funds may be conditional on accepting and excelling within a new financing modality.

New financing modalities may also complicate implementation of the Global Fund’s existing counterpart financing policy. At present, the Global Fund requires that countries meet a minimum threshold for government contributions to the national disease program and that the contribution increase over time.⁴⁷ But under a next generation financing model, the total amount of Global Fund financing could vary substantially—and Global Fund results-based payments may effectively reimburse the government for costs that

they had previously incurred. To move forward with these new models on a broader scale, the Global Fund may need to modify its counterpart financing modality to account for this additional complexity.

Finally, next generation financing models should be considered in designing and implementing the Global Fund’s forthcoming sustainability framework and differentiation initiative. These new models offer considerable advantages for sustainability and transition that should be reflected in the document. In particular, they loosen Global Fund micromanagement of implementation arrangements, they incentivize more efficient service delivery, they serve as a pull mechanism to strengthen data and management systems, and they push countries to prefinance at least a portion of service delivery costs through the national budget process. Similarly, the differentiation initiative should reflect next generation grant models as one of several tools to better tailor grant design to country and disease context.

4. Reflect the needs and requirements of next generation grants in the guidance and terms of reference given to key Global Fund bodies, including the Technical Review Panel, Country Coordinating Mechanisms, and operational divisions within the Secretariat.

The adoption of next generation grant models will also have important implications for several key Global Fund bodies, particularly the Technical Review Panel, the Country Coordinating Mechanism, and several operational divisions within the Secretariat. As a result, each body should receive new guidance and terms of reference to reflect the needs and requirements of next generation grants.

Technical Review Panel

Historically, the Technical Review Panel has focused on the technical soundness of the programmatic approach, including the choice of commodities and delivery strategy. For a next generation grant, the Technical Review Panel should instead focus on whether the selected indicators are technically appropriate given the epidemiological context and disease response challenges for any given country, whether the indicators meet the criteria for contractibility, whether the payment amount and strategy are justified, and whether the

verification strategy is sufficiently robust to ensure a sound basis of payment.

Country Coordinating Mechanism

The role of the Country Coordinating Mechanism will not necessarily change under next generation grant models, but it will need to focus on slightly different considerations. As is currently the case, the Country Coordinating Mechanism would still be responsible for nominating the Principal Recipient; however, it would now need to ensure that its selected Principal Recipient was able to perform to an acceptable standard under the parameters of a next generation grant agreement. Likewise, the Country Coordinating Mechanism would still be responsible for the creation and negotiation of the grant proposal. However, the proposal and negotiations would focus on the indicators, payment schedule, and verification strategy, not inputs and activities. Finally, the Country Coordinating Mechanism would need to continue its oversight of grant implementation and actively support the Principal Recipient's success, for example, by connecting the Principal Recipient with technical assistance from the WHO and other technical partners.

Secretariat Operational Divisions

Many Secretariat operational divisions would need new guidance and policies to appropriately manage next generation grants. These include Country Teams, Legal and Compliance, Finance, and Risk Management. To flesh out the implications for each operational division, the Global Fund could form a smaller internal working group to consider the relevant issues and ensure that guidance throughout the Global Fund is comprehensive and coherent.

5. Assure Global Fund and Principal Recipient access to needed expertise and resources to design and operationalize next generation grants, with particular attention to performance verification.

To move toward next generation grant models, the Global Fund will need to access new sources and types of expertise—for example, in health economics, incentive design, and rigorous, statistically sound

verification. The Global Fund may need to train existing staff or add new staff positions; it may also require new partnerships to access such expertise externally—for example, from the World Bank and Inter-American Development Bank, two institutions with substantial expertise in results-based aid programs. The current role of the Local Fund Agent also merits reconsideration. The Local Fund Agent may not be the appropriate body to conduct rigorous verification work, and its role as a financial overseer may become less relevant if next generation financing models are applied at scale.

Verification will require special attention. It is possible that more rigorous performance verification—as described in this report—is itself an important reform that can create virtuous incentives for better performance, independent of whether funding is tied directly to that performance. Careful attention to building capacity in this area is urged given that it is a different approach—and plays a different role for grant management—from the Global Fund's traditional support to country-led monitoring and evaluation.

In addition, the Global Fund could benefit from creating a resource hub for learning and sharing, both across the organization and externally with its Principal Recipients and other donor agencies. A core functional team could take on this role, helping document and share experiences to ensure that institutional memory, learnings, and expertise are available to Fund Portfolio Managers and their teams.

6. Revise Key Performance Indicators to accommodate differences in the management and evaluation of next generation grants.

The Global Fund's Board has approved a set of Key Performance Indicators to track the Global Fund's overall performance. However, some of the current Key Performance Indicators—particularly those covering operational aspects of the Global Fund's current funding model—may need to be revised in light of next generation financing models. Specifically, the following Key Performance Indicators may be incompatible with new financing models and should be revised:

- Key Performance Indicator 7a: Time from final Concept Note submission to first disbursement;
- Key Performance Indicator 11: Percentage of forecast grant expenses made to schedule.

7. Evolve financial management policies to accommodate less predictable cash flow and reduce restrictions on the use of funds.

Done right, next generation financing models will introduce uncertainty into cash flow projections.ⁱ To adopt these models, the Global Fund will need to consider the degree to which it can handle cash flow uncertainty. It will also need a mechanism to reallocate savings (for example, undisbursed funds) across grants. For example, it will need to decide (before grant signature) whether a Principal Recipient could recoup those funds in the next year in case of exemplary

performance or whether they would simply be reallocated to the general or supplemental fund.

More generally, the Global Fund would also need to define guidelines for use and oversight of results-based payments. Ideally, those payments would have only minimal strings attached—for example, a requirement that they be spent within the health sector and in compliance with a negative list of prohibited items (for example, arms or drugs), as is common among the multilateral development banks. Those funds would not be subject to regular audit, but the Global Fund could nonetheless reserve the right to conduct an audit at any time.

i. Since the uncertainty on any one grant will be larger than the uncertainty across many results-based grants, the Global Fund could improve the predictability of its budgeting process by funding the results-based grant components of many grants from a single budget line.

Notes

1. See Center for Global Development (2013).
2. The Global Fund, “Grant Portfolio”; <http://www.theglobalfund.org/en/portfolio/>.
3. The Global Fund, “Overview”; <http://www.theglobalfund.org/en/overview/>.
4. The Global Fund, “Grant Portfolio”; <http://www.theglobalfund.org/en/portfolio/>.
5. Glassman, Fan, and Over 2013.
6. UNAIDS 2014.
7. Birdsall and Savedoff 2010, p. 9.
8. Wren-Lewis (forthcoming); see also Estache and Wren-Lewis (2009).
9. Perakis and Savedoff 2015.
10. Fan and others 2013.
11. Macro International 2008, p. 10.
12. Macro International 2009, p. 31.
13. Perakis and Savedoff 2015.
14. Gertler and Giovagnoli 2014.
15. Basinga and others 2011; Gertler and Vermeersch 2013; Jack, 2005.
16. HRITF 2014.
17. HRITF 2014.
18. HRITF 2014.
19. Mohanan and others 2014.
20. Perakis and Savedoff 2015.
21. Luo and others 2014.
22. Cashin and others 2014.
23. Cashin and others 2014.
24. Nyweide and others 2015.
25. See, for example, figure 2 of International Evaluation Group (2012).
26. This section draws in part on Padian and Geng (2015).
27. Padian and Geng forthcoming.
28. Padian and Geng, forthcoming.
29. Padian and Geng, forthcoming.
30. Center for Global Development 2013.
31. This section draws from forthcoming background papers authored by Liam Wren-Lewis and Han Ye. It is informed by discussions at a technical workshop held at the University of California, Berkeley, in April 2015, attended by Arnab Acharya, Sebastian Bauhoff, Stefano Bertozzi, Dave Burrows, Jed Friedman, Elvin Geng, Paul Gertler, Kate Harris, William Jack, Mead Over, Nancy Padian, H el ene Roger, Nalinee Sangruee, William Savedoff, Rachel Silverman, Liam Wren-Lewis, and Han Ye.
32. This section describes adaptations of the Vogelsang-Finsinger and retail price index-X mechanisms. For an exposition, see Ye (forthcoming).
33. This section is heavily informed by Burrows and McCallum (2015).
34. The Global Fund 2012, p. 2.
35. Burrows and McCallum 2015, p. 12.
36. Burrows and McCallum 2015.
37. Burrows and McCallum 2015.
38. Burrows and McCallum 2015.
39. Burrows and McCallum 2015.
40. Glassman, Fan, and Over 2013.
41. Baser and others 2009.
42. Goldman and others 2011.
43. Dillon 2012.
44. Birdsall and Savedoff 2010.
45. Msyamboza and others, 2009.
46. The authors are indebted to Dave Burrows for his assistance in building out this case study.
47. The Global Fund, “Counterpart Financing”; <http://www.theglobalfund.org/en/fundingmodel/counterpartfinancingwtp/>.

Bibliography

- Akerlof, G.A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84 (3): 488–500.
- Arrow, K.J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53 (5): 941–73.
- Baser, O., Z. Fan, J.B. Dimick, D.O. Staiger, and J.D. Birkmeyer. 2009. "Outlier Payments for Cardiac Surgery and Hospital Quality." *Health Affairs* 28 (4): 1154–60.
- Basinga, P., P.J. Gertler, A. Binagwaho, A.L. Soucat, J. Sturdy, and C.M. Vermeersch. 2011. "Effect on Maternal and Child Health Services in Rwanda of Payment to Primary Health-care Providers for Performance: An Impact Evaluation." *The Lancet* 377 (9775): 1421–28.
- Birdsall, N., and W.D. Savedoff. 2010. *Cash on Delivery: A New Approach to Foreign Aid*. Washington, D.C.: Center for Global Development.
- Burrows, D., and L. McCallum. 2015. "Potential Application of Results-Based Financing to Key Populations HIV programming in Eastern Europe/Central Asia and Latin America/Caribbean." Background Paper prepared for the Working Group on Next Generation Financing Models in Global Health. Center for Global Development, Washington, DC.
- Cashin, C., Y.-L. Chi, P.C. Smith, M. Borowitz, and S. Thomson, eds. 2014. *Paying for Performance in Health Care: Implications for Health System Performance and Accountability*. New York: Open University Press.
- Center for Global Development. 2013. "Investing in Social Outcomes: Development Impact Bonds. The Report of the Development Impact Bond Working Group." Center for Global Development, Washington, D.C.
- Centers for Medicare & Medicaid Services. 2014. "Accountable Care Organizations: What Providers Need to Know." https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/ACO_Providers_Factsheet_ICN907406.pdf.
- Dillon, B. 2012. "Using Mobile Phones to Collect Panel Data in Developing Countries." *Journal of International Development* 2(4): 518–27.
- Estache, A., and L. Wren-Lewis. 2009. "Toward a Theory of Regulation for Developing Countries: Following Jean-Jacques Laffont's Lead." *Journal of Economic Literature*, 47(3): 729–70.
- Fan, V.Y., D. Duran, R. Silverman, and A. Glassman. 2013. "Performance-based Financing at the Global Fund to Fight AIDS, Tuberculosis and Malaria: An Analysis of Grant Ratings and Funding, 2003–12." *Lancet Global Health* 1(3): e161–68.
- Gertler, P., and C. Vermeersch. 2013. "Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes." NBER Working Paper No. 19046, National Bureau of Economic Research, Cambridge, MA.
- Gertler, P., and P. Giovagnol. 2014. "Rewarding Provider Performance to Enable a Healthy Start to Life: Evidence from Argentina's Plan Nacer." World Bank Policy Research Working Paper No. 6884, World Bank, Washington, D.C.
- Glassman, A., V. Fan, and M. Over. 2013. *More Health for the Money: Putting Incentives to Work for the Global Fund and Its Partners*. Working Group on Value for Money in Global Health, Center for Global Development, Washington, D.C.
- The Global Fund. 2012. *The Global Fund Strategy 2012–2016: Investing for Impact*. Geneva: The Global Fund.
- . 2013. "The Global Fund Corporate Key Performance Indicator Framework for 2014–2016." The Global Fund, Geneva.
- . 2014a. "2015 Corporate Work Plan." Thirty-Second Board Meeting. Geneva: The Global Fund.
- . 2014b. "Proposed Changes to The Global Fund Risk and Assurance Processes." Thirty-Second Board Meeting. Geneva: The Global Fund.
- Goldman, T.R., K.M. King, M.K. Sparrow, T. Agres, A. Schwartz, and S. Dentzer. 2011. "Health Policy Brief: Eliminating Fraud and Abuse." *Health Affairs*. http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=72.

- Green, R. 1997. "Has Price Cap Regulation of U.K. Utilities Been a Success?" Note No. 132, Public Policy for the Private Sector, World Bank Group, Washington, D.C.
- Hallett, T.B., and M. Over. 2010. "How to Pay 'Cash on Delivery' for HIV Infections Averted: Two Measurement Approaches and Ten Payout Schedules." Working Paper 210. Center for Global Development, Washington, D.C.
- HRITR (Health Results Innovation Trust Fund). 2014. "RBF: A Smarter Approach to Delivering More and Better Reproductive, Maternal, Newborn, and Child Health Services. 2014 RBF Progress Report." World Bank Group, Washington, D.C.
- Ibarrarán, P. 2015. "Target Setting within Resource Constraints." Presented at the "Working Group on Next Generation Financing Models," Veyrier du Lac, January 22.
- International Evaluation Group. 2012. *Designing a Results Framework for Achieving Results: A How-To Guide*. Washington, D.C.: World Bank Group.
- Jack, W. 2005. "Purchasing Health Care Services from Providers with Unknown Altruism." *Journal of Health Economics* 24: 73–93.
- Luo, R., G. Miller, S. Sylvia, S. Rozelle, and M. Vera-Hernandez. 2014. "Can Bureaucrats Be Paid like CEOs? School Principal Incentives for Anemia Reduction in rural China." Presented at "Making Impact Evaluation Matters," Manila, September 14. http://www.3ieimpact.org/media/filer_public/2014/09/18/32education_in_asia__luo.pdf.
- Macro International. 2009. *The Five-Year Evaluation of the Global Fund to Fight AIDS, Tuberculosis, and Malaria. Synthesis of Study Areas 1, 2 and 3*. Calverton, Md.: Macro International.
- Mohanan, M., S. Bauhoff, G.L. La Forgia, K. Singer Babiarz, K. Singh, and G. Miller. 2014. "Effect of Chiranjeevi Yojana on Institutional Deliveries and Neonatal and Maternal Outcomes in Gujarat, India: a difference-in-differences analysis." *Bulletin of the World Health Organization* 92 (2): 187–94.
- Msyamboza K.P., E.J. Savage, P.N. Kazembe, S. Gies, G. Kalanda, U. D'Alessandro, and B.J. Brabin. 2009. Community-based Distribution of Sulfadoxine-pyrimethamine for Intermittent Preventive Treatment of Malaria during Pregnancy Improved Coverage but Reduced Antenatal Attendance in Southern Malawi." *Tropical Medicine and International Health* 14(2): 183–89.
- Nyweide, D.J., W. Lee, T.T. Cuerdo, H.H. Pham, M. Cox, R. Rajkumar, and P.H. Conway. 2015. "Association of Pioneer Accountable Care Organizations vs Traditional Medicare Fee for Service with Spending, Utilization, and Patient Experience." *JAMA* 313 (21): 2152–61.
- Over, M. 2011. *Achieving an AIDS Transition: Preventing Infections to Save Lives*. Washington, D.C.: Center for Global Development.
- Over, M., and M. Ravallion. 2015. "Recognizing and Rewarding the Best Development Professionals. Global Development: Views from the Center." Center for Global Development. <http://www.cgdev.org/blog/recognizing-and-rewarding-best-development-professionals>.
- Padian, N., and E. Geng. (forthcoming). "Recommended Indicators for Performance-based Contracting." Background paper prepared for the Working Group on Next Generation Financing Models in Global Health. Center for Global Development, Washington, DC.
- Perakis, R., and W. Savedoff. 2015. "Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies." Policy Paper 052, Center for Global Development, Washington, D.C.
- Pollitt, M.G. 1999. "A Survey of the Liberalization of Public Enterprises in the UK since 1979." University of Cambridge, Cambridge.
- Sandefur, J., and A. Glassman. 2013. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." Working paper 373, Center for Global Development, Washington, D.C.
- UNAIDS (United Nations Joint Program on HIV/AIDS). 2014. "90–90–90: An Ambitious Treatment Target to Help End the AIDS Epidemic." United Nations Joint Program on HIV/AIDS, Geneva.
- Wren-Lewis, L. (forthcoming). "Designing Contracts for the Global Fund: Lessons from the Theory of Incentives." Center for Global Development, Washington, D.C.
- Ye, H. (forthcoming). "Global Health Donors Viewed as Regulators of Monopolistic Service Providers: Lessons from Regulatory Literature." Center for Global Development, Washington, D.C.



The Global Fund to Fight AIDS, Tuberculosis and Malaria is one of the world's largest multilateral health funders, disbursing \$3–\$4 billion a year across 100-plus countries. Many of these countries rely on Global Fund monies to finance their disease responses. For their citizens, the efficient and effective use of Global Fund monies can be the difference between life and death.

One way the Global Fund can advance its strategic interests and accelerate the impact of its investments would be to tie grant payments to achieved and verified results. This report, the result of a Center for Global Development Working Group on Next Generation Financing Models in Global Health, addresses how to implement such payment models. It offers concrete steps that the Global Fund will need to take to put the new models into place and maximize their potential benefits.



ISBN 978-1-933286-94-5