

5

A Learning Agenda

What elements of performance incentive programs lead to success? What pitfalls can be avoided? When do performance-based programs generate more bang for the buck than other approaches? What tools are needed to help governments and nongovernmental organizations (NGOs) put performance-based financing in place? Moving beyond the monitoring and evaluation that should be a part of any performance-based program (chapter 4), here we propose an agenda for learning that extends beyond any individual country or program. It is about developing knowledge and tools that can be used widely, rigorously measuring and understanding what works across settings, and creating an ongoing way to share and learn among those who are implementing and studying performance-based programs.

Filling the Toolbox

Our look at several real-world cases suggests that three types of new tools would be particularly useful in strengthening future programs.

Standardized Assessments

To date, most performance-based initiatives in the developing world have focused on either overcoming household- or patient-related barriers or changing provider

reimbursement to affect the behavior of individual health workers, managers, and the systems to deliver services. Few have tackled both at the same time. The choice—demand, supply, or both—has rarely been driven by an explicit assessment of all barriers; instead, it has usually been based on the starting premise and mandate of the designers. Even more neglected are the actions required by other actors, such as other sectors or community-based initiatives, to support the effective use of essential interventions. A comprehensive assessment of the causes of poor performance would be a sound first step toward designing performance-based programs.

Such an assessment might include the following:

- Methods to measure the extent of problems with use, efficiency, and quality of service,

- Diagnostic questions, analysis, and qualitative methods to assess provider productivity and quality and to understand existing incentives and their effects on behavior,

- Analysis of household survey data to quantify household barriers to effective use of health services,

- Diagnostic questions to clarify objectives and prioritize problems that the incentives would address,

- Analytic tools, such as worksheets, to estimate the costs of implementing a performance-based program, both the near-term costs of switching and the likely recurrent costs,

- Guidelines for mapping the interest groups likely to favor or oppose a performance-based incentive approach, and

- Tools to model the potential impact of incentives, such as a simulation model to ask “what if” under various assumptions.

A related tool with some of these features has been developed, field tested, and refined for tuberculosis control programs (see Weil and others 2004). This could be adapted for other applications.

Performance Indicators

A dynamic handbook on performance indicators would address the challenge of selecting which indicators are appropriate in given circumstances. Such a volume would also bring together the best available evidence about the link between desired health outcomes and particular observable behaviors that might be changed with the introduction of incentives. This is particularly important for provider-side interventions, which often seek not only to increase the use

of key services but also to improve the quality of service delivery. Approaches to measuring responsiveness and satisfaction of households would also be included, with sample instruments to measure user satisfaction that could be implemented cost-effectively at the community or service provision level. The starting point would be to develop a set of evidence-based process and output measures related to quality, drawing from evidence-based guidelines and consumer assessment instruments.¹

The handbook could also provide guidance on methods to measure and validate performance that are feasible in developing-country contexts, including approaches to measuring and validating performance that can be implemented in countries with paper medical records and limited access to technology. Specifications needed to measure and validate results through an independent third party and systems for recipients to self-report coupled with random audits could also be included. Step-by-step guides to monitor and verify results that could be adapted to a specific context would reduce start-up costs.

Contracts and Agreements

A compendium of contracts and performance agreements detailing how payment is linked to performance would provide a useful menu of options. Included could be contracts that specify performance indicators and targets, how performance is measured, and how payment is linked to results. The contracts might also incorporate notes from the designers and implementers that explain why the specific indicators and payment approaches were chosen and the reasons for modifications and lessons learned during implementation.

The handbook and compendium of contracts could be organized as online resources, be updated at relatively low cost, and invite contributions from researchers, practitioners, and others. The online documentation could be supplemented by occasional expert panel reviews of indicators and terms of the contract.

Assessing the Impact

Financiers, governments, and policymakers at every level want to know which approaches to paying for performance have the greatest impact and when paying for performance is more effective than other approaches. They would also like to avoid the failures of other performance incentive programs and replicate the

1. Websites include www.who.int/child-adolescent-health/publications/pubIMCI.htm, ih.jhsph.edu/chr/fhacs/imci2.htm [October 2008], and www.ebmny.org/cpg.html [October 2008].

successes. To address those needs, we can examine, closely and rigorously, the effects of relevant alternative approaches, to establish a causal link between programs and outcomes. Analyzing program performance after the fact through routine program monitoring data and quasi-experimental research methods is productive, but subject to biases that are difficult to correct and therefore make it hard to draw reliable conclusions.

To measure the impact of a new program, it is best to observe the same individuals or providers in parallel situations, with and without the program, at the same time. The comparison group can be created any number of ways, including random assignment, statistical matching, and eligibility filtering. Of these, random assignment is most likely to avoid biased results, although obstacles to its use in many settings are significant. Such an approach enables us to account for other factors that may determine success of those in the program, such as whether more capable providers—those most likely to be high performers without additional financial incentives—elect to participate.

In all the cases highlighted in this book, packages of interventions were implemented simultaneously, making it hard to attribute an improvement in performance to any one of them. For example, many supply-side programs include technical assistance, increased funding, strengthened information systems, increased autonomy, and precise definition of expected results accompanied by improved monitoring. Is improved performance coming from the financial incentive, the clear expectations, the better information, or the increased autonomy? Most likely, it is attributable to a combination of factors, and the package as a package is the relevant intervention to examine, not individual elements. More information is needed, though, to disentangle the contributions of interventions that are introduced simultaneously.

The Nicaragua program faced an additional challenge because it included both supply-side and demand-side incentives. Its impact evaluation examined the impact of the package but did not disentangle the contribution of the elements. To establish the relative importance of different interventions, different treatments, such as technical assistance, funding, and technical assistance plus funding, would have to be compared against each other and a comparison group. Choices among possible comparisons to study should be based on real-world options.

Impact evaluation is more than a tool for gauging impacts at the end of a program and providing inputs into a cost-effectiveness analysis. It can also help a program to evolve. For example, in the initial phase of a pay-for-performance program, three contracts with different risk levels can be piloted. Based on the results

from an early evaluation, the most effective contract can be scaled up. Several parameters lend themselves to this kind of experimentation, including the relative effects of supply versus demand interventions, the level of rewards offered for performance, and the balance of trade-offs between access and use.

In addition, impact evaluations should collect data to allow for possible unintended perverse or positive effects. Monitoring to determine whether increased attention on rewarded services leads to neglect of other important services is critical. The impact on the distribution of health workers also requires careful assessment because the potential to earn performance awards can either exacerbate or mitigate regional differences, depending on the design. Careful monitoring and evaluation can help to identify positive spillover effects such as an increased use of services, rewarded and not, by a previously underserved population group.

On a local level, impact evaluation can be valuable in maintaining and strengthening political support. In addition, although no findings from one context can be generalized without qualification, the impact evaluation of one program does provide valuable lessons, such as a benefit that is amplified as evidence from a number of similar programs is combined. A good example on the demand side is conditional cash transfers (see chapter 6). For other demand-side interventions, however, as well as for most supply-side interventions, the evidence remains thin.

Evaluations that provide lessons for other contexts ideally would describe the market for services—number of providers, potential for competition—that existed before and the problems and constraints that the incentives were designed to address. A description of the landscape of service provision that includes the number of providers serving a given population would be useful contextual information.

Understanding the incentives that providers faced before performance incentives were introduced may enable insights into the applicability of a given experience to another context. Environments in which providers are paid a salary not linked to performance differ from those in which they are given capitation payments or fees. The new incentives interact with the existing incentives. A program that pays performance awards by holding back or reallocating money from existing budgets may affect behavior differently than a program that funds performance awards by using new funding. The same design, for example, might have desired effects when funded by infusing new money and weak effects when funded by reducing existing budgets. Although the source of funding is not likely to be an element of an impact evaluation in a specific environment, it is an element

of design and context that needs to be noted when building a global body of evidence.

When to Consider Programs

Performance incentives are one in a basket of potential approaches to improve health results. How do donors and policymakers choose? Key to this decision is whether and under what circumstances using money to buy results generates a higher return. Evaluations are needed that compare performance incentives to training and other approaches.

Another important question is whether the benefits justify the costs incurred. In addition to immediate-term benefits, such as increased use of services, performance incentives may provide benefits that will be realized only over decades, such as strengthening the capacity of delivery systems and alleviating poverty. Comparisons of alternate interventions need to attempt to capture such multiyear benefits.

Key Questions

Impact evaluations and other research efforts are most useful when they address questions that can inform future design and implementation decisions. Future research might well incorporate and examine the questions that follow, which are in no way intended as comprehensive.

—*Demand-side programs.* The magnitude of financial transfers needed to achieve health goals,² whether the costs of conditioning payment justify the benefits, whether communicating the conditions of income transfer without monitoring compliance is enough, how changes in demand affect provider actions, and whether there are unintended consequences and effective strategies to avoid negative effects.

—*Supply-side programs.* The effects and costs of providing incentives to individuals and to teams, the advantages and disadvantages of schemes that pay for each service over those that pay on the basis of population-based targets or a balanced score card, the share of provider income to be at risk and the form pay-

2. An impact evaluation is being implemented in Malawi that is examining the effect of none to significant financial payments to people who are HIV negative if they maintain their negative status after one year. People are randomly assigned at the point of their HIV test to payment regimes that vary from no payment to the equivalent of two months' average wage. In addition to providing information about whether performance incentives work, this study will provide more refined information to inform how much payment yields the desired result and whether there are diminishing returns (Rebecca Thornton, discussion, December 2007).

ment should take, the best target-setting strategy (provider baseline, absolute threshold, tournament style, or a combination), the effects of implementing a program by reallocating existing budgets and by using new funding, and how to avoid negative effects and enhance positive effects.

Creating a Network

Some of the most important knowledge needed for successful performance-based incentive programs will be gained by trial and error and be captured by those undertaking the challenge of implementing programs. A serious global learning agenda should include creating a learning network of funders, researchers, and program managers across countries.

A starting point might be an interactive website offering an “ask the expert” series, case studies, and a library of performance incentives. As incentive programs grow and lessons accumulate, the benefits of global networking to share lessons will also grow.

Likely candidates for a network approach include supply-side strategies that provide performance-based payment at the institutional level or higher. A network of payers might also be a helpful way for a range of programs to learn from one another. National programs that implement performance-based payment programs can benefit from cross-country networks to exchange lessons. For example, tuberculosis control programs that have implemented some form of either supply-side or demand-side incentives have met three times at the annual meeting of the International Union Against Tuberculosis and Lung Disease to share lessons and to reach consensus on incentive approaches that increase both case detection and treatment completion. The World Bank has organized international meetings that bring together delegations from across the globe to discuss conditional cash transfer programs. Officials and researchers from Mexico, Chile, Brazil, Argentina, and Colombia meet frequently to exchange lessons about CCT programs, and Mexico frequently hosts delegations from other countries wishing to view the program in action.

Conclusions

The learning agenda set out here is ambitious and important. It calls for new tools to help practitioners to implement performance incentive programs and to link people across the globe in a dynamic and interactive framework that allows them to share and learn from one another. It also calls for a commitment to conduct

impact evaluations that will generate the sound body of evidence needed if financiers and program implementers are to understand whether and when there is value in paying for results.

Reference

Weil, Diana, and others. 2004. "Mapping the Motivations of Stakeholders to Enable Improved Tuberculosis Control: Mapping Tool for Use in Workshops." Arlington, Va.: Management Sciences for Health and the Stop TB Partnership.

Copyright © 2009

CENTER FOR GLOBAL DEVELOPMENT

1800 Massachusetts Avenue, N.W.

Washington, D.C. 20036

www.cgdev.org

Performance Incentives for Global Health: Potential and Pitfalls may be ordered from:

BROOKINGS INSTITUTION PRESS

c/o HFS, P.O. Box 50370, Baltimore, MD 21211-4370

Tel.: 800/537-5487; 410/516-6956; Fax: 410/516-6998; Internet: www.brookings.edu

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the Center for Global Development.

Library of Congress Cataloging-in-Publication data

Eichler, Rena.

Performance incentives for global health : potential and pitfalls / Rena Eichler, Ruth Levine and the Performance-Based Incentives Working Group.

p. ; cm.

Includes bibliographical references.

Summary: "Describes the rationale for introducing incentives tied to achievement of specific health-related targets, and provides guidance about designing, implementing, and evaluating programs that provide incentives to health care providers and patients. Presents case studies that focus on recent uses of incentives addressing a range of health conditions in diverse countries"—Provided by publisher.

ISBN 978-1-933286-29-7 (pbk. : alk. paper)

1. Medical economics. 2. World health. 3. Health promotion. I. Levine, Ruth, 1959– II. Center for Global Development. Performance-Based Incentives Working Group. III. Title.

[DNLM: 1. Delivery of Health Care—economics. 2. Program Evaluation—economics. 3. Reimbursement, Incentive—economics. 4. World Health. W 84.1 P4376 2009]

RA410.5.E43 2009

338.4'73621—dc22

2009000907

9 8 7 6 5 4 3 2 1

The paper used in this publication meets minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials: ANSI Z39.48-1992.

Typeset in Adobe Garamond

Composition by Circle Graphics, Inc.

Columbia, Maryland

Printed by Versa Press

East Peoria, Illinois