National Drug Stockout Risks and the Global Fund Disbursement Process for Procurement

Jérémie Gallien

London Business School, Regent's Park, London, NW1 4SA, UK, jgallien@london.edu

Iva Rashkova

Olin Business School, Washington University, St. Louis, Missouri 63130, USA, irashkova@wustl.edu

Rifat Atun

Harvard School of Public Health, Harvard University, Boston, Massachusetts 02138, USA, ratun@hsph.harvard.edu

Prashant Yadav

University of Michigan School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA, yadavp@umich.edu

Despite substantial financial aid from international donors for procurement of health products, stockouts of life-saving drugs related to prevalent infectious diseases are still widespread in Africa. Rigorous research to understand the underlying causes of these stockouts is lacking. To this end, we study the relationship between The Global Fund to Fight AIDS, Tuberculosis and Malaria and its grant recipients. Specifically, we leverage historical fund disbursement and drug procurement data from 2002 to 2013 to build a discrete-event simulation model predicting the joint impact of procurement and grant disbursement processes on national drug availability for the Global Fund's recipient countries in Africa. This model is validated against cumulative stockout levels inferred from historical grant implementation lengths, and used to evaluate potential high-level modifications in the disbursement or procurement process. Results show the existence of significant intrinsic stockout risks in most African countries, with particularly high levels in East Africa, due to the unpredictability of fund disbursements and the frequency of grant performance monitoring performed by the Global Fund. Interventions shifting some fund disbursements upfront to protect against disbursement timing uncertainty are predicted to be more effective than others that include regional buffer stocks and bridge financing.

1. Introduction

Twenty-five years ago, almost one-third of the world's population did not have access to essential medicines (Foster et al. 2006). Major trends in global health since then include the emergence of new actors such as The Global Fund to Fight AIDS, Tuberculosis and Malaria (Global Fund), the Global Alliance on Vaccines and Immunizations (GAVI), the Bill & Melinda Gates Foundation as well as budget increases of bilateral donors such as the US and UK governments (Atun et al. 2012). This has resulted in a significant increase of international funding for health programs in low-income countries (currently US $27 billion a year, see IHME 2014). Unfortunately, communicable diseases treatable in the developed world remain widespread: HIV/AIDS remains the leading cause of adult death in Africa with an estimated 23 million people living with HIV at the end of 2011 and 2.5 million new infections per year (United Nations 2013); malaria and tuberculosis combined led to over 2 million deaths in 2011, again mostly in Africa (WHO 2013a, b).

Established in 2002, The Global Fund is currently the world's largest external financier of HIV, Tuberculosis and Malaria programs. Funded by countries such as the United States (29% of total paid to October 2012), France (13%), the United Kingdom (9%), and Germany (7%); private foundations such as the Bill and Melinda Gates Foundation (5%); and corporations such as Chevron (0.2%), by October 2012 it had committed US $22.9 billion to prevention, treatment, and care in 151 low and medium-income countries. This includes United States $9.2 billion (around 40%) for procuring medicines and health products, and United States $9 billion (around 38%) for strengthening health systems (Global Fund 2012a).

The Global Fund was the pioneer global health organization serving as a financier only without a direct role in health program implementation. To raise money from its donors, the Global Fund promotes its specific performance-based financing model: disbursements to grant recipients are conditional on past grant performance, which involves fund usage transparency and achievement of result targets predefined by the grant recipient (e.g., number of patients treated, number of doctors trained, number of facilities opened, see Center for Global Development 2013). Although other organizations such as GAVI and the World Bank have recently adopted performance-based funding for some of their activities, the Global Fund constitutes the first and largest global implementation of this innovative funding model to date. As a result, the Global Fund experience presents a unique opportunity to identify lessons about performance-based funding that may be relevant to many other organizations.

Indeed, despite the positive impact of Global Fund-supported programs (Brugha et al. 2004), stockouts of health products at health facility (peripheral) and national level (e.g., central warehouse) have been widespread in countries receiving Global Fund financing, particularly in Africa (Oliynyk 2011, PLoS Medicine Editors 2009, Yu et al. 2008): in a 2009 survey, 9 out of 14 surveyed African countries reported stockout of at least one type of medicine related to Global Fund grants within the last year, four reported stockouts of two or more types, and all reported at least one near-stockout situation (Global Fund 2009). Stockouts cause treatment interruptions, loss of confidence in health systems and providers, increased risks of drug resistance and adverse effects on disease epidemiology. Consequently, stockouts lead to increased morbidity and mortality for a large number of patients receiving treatment for AIDS, Tuberculosis and Malaria and pose a major challenge to public health (Hawkes 2011, Levine et al. 2008, WHO 2004). In general, stockouts of medicines in Africa have been attributed to procurement delays (ALMA 2011), fund disbursement delays (Celasun and Walliser 2007, Lane and Glassman 2008), and insufficient visibility of stock levels in peripheral health facilities (Shretta and Yadav 2012). While the existing literature does include rich contextual observations of stockouts, rigorous quantitative research on their causes is lacking.

This paper is an empirical study of the relationship between national stockout risks for health products purchased with Global Fund grants in Africa and the process used by the Global Fund for performance monitoring and procurement fund disbursements. Specifically, we leverage publicly available historical data for Global Fund grants in Africa between 2002 and 2013 to build a discrete-event inventory simulation model predicting the joint impact of procurement and grant disbursement processes on national drug availability in recipient countries. This model and some of its assumptions (e.g., single product, deterministic demand, and single funding source) may appear simplistic at first glance. Notably however, we are able to validate its predictive accuracy against cumulative stockouts inferred from historical grant implementation lengths, the primary output measure of interest. These validation results provide some justification of our use of this model to address the following main questions:

1. What is the impact of the grant recipient performance monitoring frequency (i.e., the scheduled frequency of grant disbursements and reporting and monitoring activities) used by the Global Fund on the stockouts experienced by receiving countries? The results discussed in section 4.1 suggest that grants with higher reporting frequency exhibit substantially higher stockout risks. Hence, over the first 11 years of the Global Fund, there was a clear effective trade-off between the extent of its performance monitoring activities and the effectiveness of the procurement funds it disbursed.

2. Are there some geographic patterns affecting the risks of stockouts experienced by Global Fund grant recipients? We find that the African regions used internally by the Global Fund for organizational purposes constitute a substantial driver of stockout risks variability, with grant recipients in East Africa facing significantly higher stockout risks than in other regions. This suggests that a substantial fraction of the stockouts facing grant recipients are driven by organizational features and specific processes used by the Global Fund, as opposed to underlying risk factors associated with these recipients (see section 4.2).

3. What is the potential impact on stockout risks of various process modifications considered by the Global Fund? We find that front loading of disbursement schedules has the potential to reduce expected stockouts much more significantly than regional buffer stocks or bridge financing (see section 4.3).

By exemplifying the application of standard operations management research methods to investigate global health challenges (Garnett et al. 2011, Kraiselburd and Yadav 2011), this work may inform the policies of the Global Fund, but also other international financing institutions using performance-based financing. This study also presents contextual information and delineates research questions that may be

useful to other researchers interested in global health operations.

The remainder of this study is organized as follows. After a review of existing related work in section 2, we discuss the definition, estimation and validation of our empirical inventory model in section 3. Section 4 presents the experiments performed with that model and their results. Concluding remarks in section 5 include a summary of our findings and their implications as well as a discussion of future research opportunities. In the remainder of this introduction, we provide additional background on the Global Fund's funding process (in section 1.1) and discuss various potential or actual interventions related to that process (in section 1.2).

1.1. The Global Fund's Grant Funding Process until 2013

Following funding round announcements by the Global Fund roughly once a year, nominated organizations (governments, NGOs, or private sector institutions) called *principal recipients* (PR) submit proposals for Global Fund financing for disease-specific programs. Subsequent approval by the Global Fund of a total program budget sets out a disbursement schedule of successive reporting periods for the awarded grants, each typically 90 or 180 days (minimum 90, maximum 360 days). The reporting frequency used for each program is determined by the Global Fund's perception of the risks associated with its implementation. Implementation risks may stem, for example, from overspending or lack of respect of budget lines, lack of suitable accounting software and procedure, excessive use of cash payments, absence of supporting documentation for expenditures, inadequate storage and distribution of pharmaceuticals, lack of transparent procedures to select or monitor subcontractors and data quality problems.[1]

After each period, PRs submit a progress report and fund disbursement request for the next period that must be consistent with the needs defined in the initial proposal. The first disbursement includes an additional cash buffer of 3 months, and similar buffers may be subsequently approved by the Global Fund (Global Fund 2012b).

In order to coordinate various aspects of the relationship with PRs, the Global Fund employs fund portfolio managers who each focus on a couple of countries and are organized in regional teams (e.g., Africa is divided in four regions). In addition, the Global Fund contracts local fund agents (academic institutions, private management consulting firms) to audit and assess programs on their behalf. Based on these agent recommendations, the Global Fund may issue program evaluation scores including:

- A—meeting or exceeding performance expectations;
- B1—adequate performance;
- B2—inadequate performance but with demonstrated potential; and
- C—unacceptably poor performance; may be discontinued.

Historically, the first 2 years of a grant were called Phase I, which recipients could often extend by a few months through specific ad-hoc requests. To ensure more predictable long-term funding beyond that first phase, recipients could then submit funding continuation applications for another 3 years called Phase II. Out of the 461 grants whose Phase I ended during the period of study, 325 (70%) were approved for Phase II funding. Formal evaluations by the Global Fund during Phase I have been far less systematic than during Phase II.

In principle, based on these scores, the Global Fund will determine its response to disbursement requests in each period. The procedure is repeated every period, with the most recent evaluation score being from the preceding period. Disbursement delays are common, and may result from missing documentation, PRs not completing performance-related preconditions identified by fund agents, or resource constraints affecting either the Global Fund or PRs. Because grants occasionally get discontinued due to poor performance and public financing is distrusted in low-income countries, disbursement completion is nearly always required before associated procurement orders can be placed from vendors. Therefore, disbursement delays can prompt emergency searches for alternative funding sources and/or affect the continuity of the local drug supply (Brugha et al. 2004).

1.2. Process Modifications Considered by the Global Fund

To reduce stockouts of medicines at the national level, several interventions related to the Global Fund financing and procurement processes have been considered. They include Pledge Guarantee for Health, a bridge financing scheme developed by the United Nations Foundation to provide funds for the period between grant approval and disbursement (UNDP 2011), which was used for the first time in the field in 2013, and international or regional buffer stocks designed to reduce procurement lead times (PLTs) (Global Fund 2011), which were tested for the first time in the field in late 2012.

In late 2012, the Global Fund announced an intention to completely redesign its legacy funding process (Global Fund 2013). While the core principles, methods for investment project selection and financing allocation are already in place, important operational

features related to procurement and disbursement are still not unified and public. The present study, which analyzes the largest currently available dataset on performance-based financing, thus informs the operational aspects of this initiative.

2. Related Literature

Existing quantitative studies of Global Fund grant operations have examined the factors influencing grant evaluation scores (Radalet and Siddiqi 2007) and cumulative disbursements (Cohen et al. 2008, Lu et al. 2006). Fan et al. (2013) have recently argued that current incentive mechanisms are not adequate as performance ratings, are not replicable by external observers, and not sufficiently connected with actual funding decisions. Our work extends this stream of research by characterizing the factors affecting Global Fund disbursement and procurement lead-times, and by quantifying the link between these lead-times and the risks of national stockouts of health products faced by Global Fund grant recipients in Africa.

There is also a recent body of work on operational issues related to donor funding for global health that is relevant to large-scale subsidy programs. In particular, Tougher et al. (2012) provide an empirical study of the Affordable Medicines Facility for malaria commodities program (AMFm) showing that subsidies combined with supporting interventions can rapidly improve availability, price and market share of quality-assured artemisinin-based combination therapies. Theoretical models of subsidies include Taylor and Xiao (2014), which consider the effectiveness of sales vs. purchase subsidies in improving the availability of malaria drugs, and show that the donor should only subsidize purchases and not sales; Levi et al. (2016) analyze an optimization model showing that uniform subsidies to competing manufacturers maximize consumption under some assumptions. In contrast, our work focuses on grants that are used for the full funding of procurement activities by grant recipients (as opposed to subsidies), which is the traditional and predominant funding channel used by the Global Fund.

Several papers in the broader operations management literature also consider the implications of uncertain lead times (Kouvelis and Li 2008, Song 1994, Song et al. 2010, Wang and Tomlin 2009) and financing (Buzacott and Zhang 2004, Chao et al. 2008, Gong et al. 2014) on inventory systems. Most relevant in this body of work is arguably the theoretical analysis by Natarajan and Swaminathan (2014), which characterizes the optimal procurement policy for a health product in the presence of funding uncertainty over a finite time horizon. While their mathematical model is closely related to our work, our intended

contribution is an empirical one that focuses on the Global Fund and performance-based funding. This contextual focus is motivated by the dominant role currently played by the Global Fund in the global financing of procurement of health products relative to other agencies such as UNICEF that primarily support in-country implementations of health programs and use more traditional fund transfer mechanisms.

This has important modeling implications, because under the performance-based funding mechanism used by the Global Fund, grant recipients submitting fund disbursement requests are required to document the satisfactory use of funds previously disbursed during past grant review periods, consistent with the initial grant agreement (see section 1.1). Therefore, Global Fund grant recipients have strong incentives to commit funds quickly after their disbursements. This can be verified empirically from our dataset, where 79% of the 3027 procurement orders funded by Global Fund grants between 2002 and 2012 were placed in the 2 weeks preceding and following a fund disbursement. This occurred even though the Global Fund grants we consider involved disbursement inter-arrival times (DITs) of 3–6 months (delays of up to 2 weeks between disbursement approval notification and actual fund transfer are observed). Given the various steps involved in public procurement processes and related data entry issues, it is also possible that many of the remaining 21% of orders were in fact committed in the days following a fund disbursement. This justifies our model assumptions that procurement orders are placed immediately after fund disbursements, and that the procurement policy of Global Fund recipients (i.e., order timing and quantity decisions) is entirely determined by the disbursement schedule.

In contrast, Natarajan and Swaminathan (2014) derive the optimal inventory policy for a more traditional and less constrained theoretical procurement model involving inventory holding costs and interest income for unused funds, and where there is no endogenous relationship between the use of funds by recipients and the timing of future disbursements. Hence, the focus of our work is on the empirical link between stockout risks and the grant-recipient interaction process, rather than determining an optimal procurement policy. Notably both papers establish, in their respective motivating contexts, that uncertainty in disbursement timing has a substantial negative impact on service levels.

Finally, our work includes a case study on the operations of a major global health organization, and an empirical analysis of related data resulting in validated distributional forecasts of PLTs for several important categories of health products in Africa. Other references providing contextual information

and data about global health supply chains include Yadav (2007), which discusses long and unpredictable PLTs for essential commodities in Zambia and their relation to drug stockouts; and Gallien et al. (2016), which contains a detailed case study of the public distribution of pharmaceuticals in Zambia and presents related datasets and a validated simulation model.

3. Simulation Model

Our simulation model is designed to characterize the empirical relationship between actual disbursement lead-times linked to the Global Fund performance monitoring process, actual procurement lead-times for health products, and the risks of national stock-outs in African countries receiving Global Fund grants. In addition, we want to evaluate the relative effectiveness of various possible interventions for reducing these stockout risks (see section 1.2).

We emphasize that our objective is therefore not to develop realistic predictions of inventory levels of actual products in specific facilities of recipient countries at any point in time. Such an objective would likely imply a considerably more complex model than is formulated here, and require more detailed data than was available to us for this study. More importantly, such a detailed tactical model would need to capture many more idiosyncratic aspects associated with any specific country setting, and thus likely fail to support our intended examination of the Global Fund's procurement and funding processes across many countries in Africa. In summary, the model to be presented here attempts to combine empirical grounding and validated predictions with a broad and strategic policy perspective.

In the remainder of this section, we first provide in section 3.1 a precise definition of our model structure. We then describe the data used (section 3.2.1) and the methods followed for estimating key model input data, including PLTs (section 3.2.2), disbursement interarrival times (section 3.2.3) and grant ratings (section 3.2.4). Finally, we discuss the results of our model validation experiments in section 3.3. Figure 1 provides an overall schematic methodology overview, and we also refer the reader to B in the online supplement for a more detailed discussion of our model assumptions than is provided here.

3.1. Model Structure Definition

Our discrete-event model simulates the inventory level I_t^{ijmlp} on day t of a single health product m procured to a central location of a country j by principal recipient i with a Global Fund grant in phase $p \in \{$Phase I, Phase II$\}$ with reporting period $l \in \{$90 days, 180 days$\}$. In many countries, this central location would correspond to the national

warehouse where public procurement orders are delivered before that inventory is shipped to patient-facing health facilities. While many principal recipients, such as ministries of health, operate in a single country (so that j is entirely determined by i), others such as the United Nations Development Programme operate in a number of African countries. The model is instantiated for 130 principal recipients in 53 African countries and the five types of health products m procured with Global Fund grants, where

$$m \in \left\{ \begin{array}{l} \text{anti-malarial, anti-tuberculosis, anti-retroviral,} \\ \text{malaria prevention and HIV prevention} \end{array} \right\}$$

$$(m-1)$$

Demand is assumed to deplete available inventory at a constant and deterministic rate, normalized to 1 per day. Our model does not capture potential changes in health product prices, so inventory levels and disbursement amounts are both measured in duration of demand coverage. We define demand occurring when there is no inventory as a *stockout* and record it as lost.

Inventory is replenished by deliveries from suppliers, which are affected by their PLTs as well as the timing and amount of disbursements by the Global Fund, which are affected by the ratings obtained during the previous reporting period. The remainder of this subsection defines the deterministic inputs, probabilistic inputs and dynamics of this replenishment model.

3.1.1. Deterministic Model Inputs. The baseline initial inventory available at the time origin I_0^{ijmlp} is set to 180 days or 6 months of demand, because this is the recommended inventory level stated in several existing guidelines for preventing stockouts (Ministry of Health, Uganda 2012, Global Fund 2006). For sensitivity analysis, it is varied between 0 and 9 months in increments of 3 months:

Baseline: $I_0^{ijmlp} = 180$ days.

Sensitivity: $I_0^{ijmlp} \in \{0, 90, 180, 270\}$ days.

$$(\text{Input} - \text{I} - 2)$$

The initial grant rating R_0^i is set to the most frequent rating in the historical disbursement database for the principal recipient considered (see data description in section 3.2.1.).

The total budget disbursed is set to 3×365 days or 3 years of demand (typical length of Phase II). This total grant budget is disbursed in several installments over the grant lifecycle. The nominal amount f of each disbursement is set to the amount necessary to cover demand for one scheduled grant period of length l (90

Figure 1   Diagram of Study Methodology [Color figure can be viewed at wileyonlinelibrary.com]
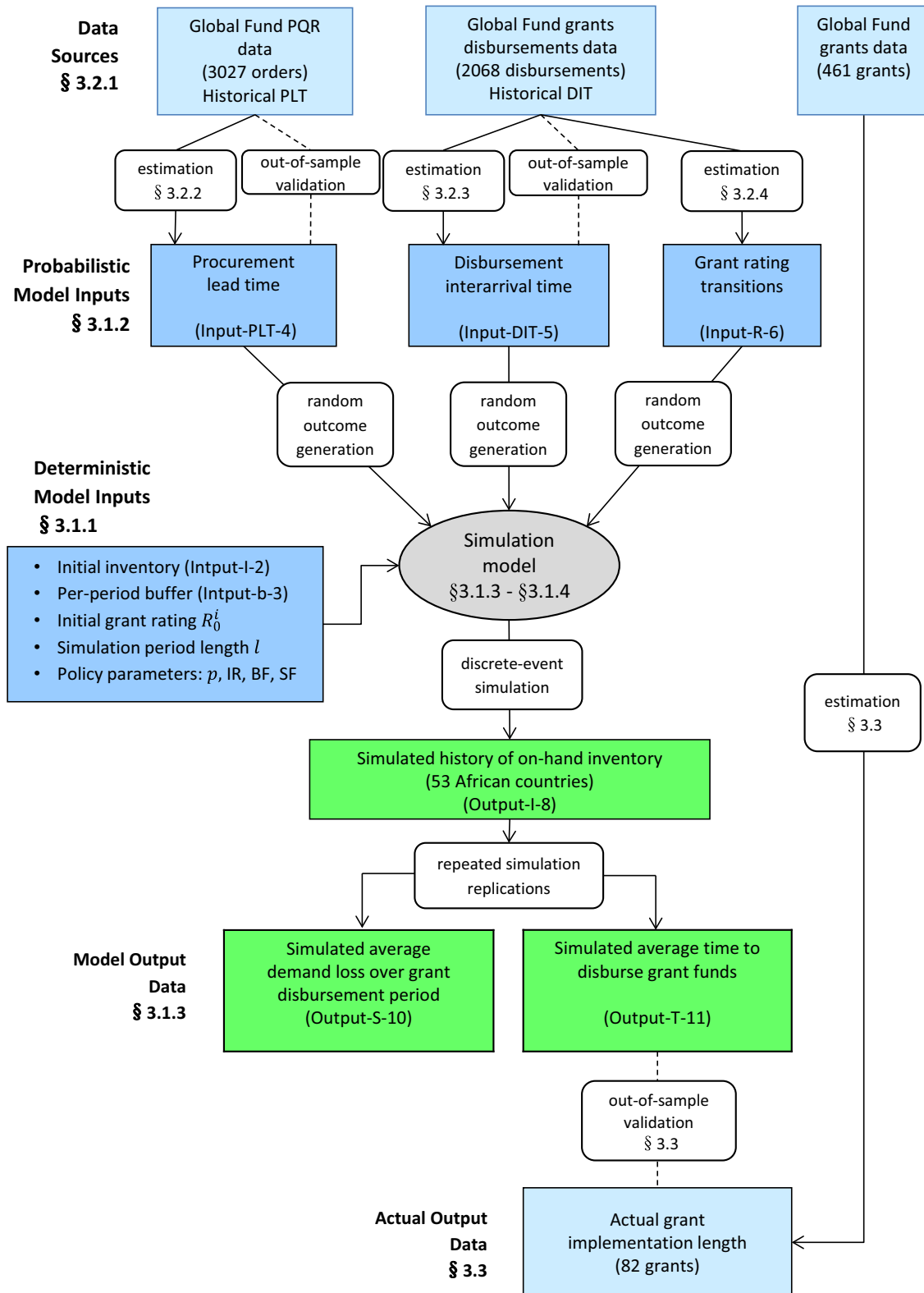
Notes: Represents data sources and methodological steps with associated input and output variables involved in the study. Location references indicate the sections in the study where an associated object is discussed. The solid line arrows indicate a methodological step linking some designated input and output. Dashed line indicate a comparison between two data sources as part of a validation step.

or 180 days) plus a *cash buffer b* expressed as a fraction (see section 1.1), so that $f = (1 + b) \times l$. The baseline cash buffer is set at $b = 10\%$ as suggested by the Global Fund (Global Fund 2012b). For sensitivity analysis, it is varied between -20% and 100% in increments of 10%:

Baseline: $b = 10\%$.
Sensitivity: $b \in \{-20\%, -10\%, 0\%, \ldots, 100\%\}$.

$$(\text{Input} - b - 3)$$

All disbursements are equal to the nominal amount f except when the remaining budget is smaller than f, at which point the last disbursement is set to the remaining budget. The total number of disbursements is thus $\lceil (3 \times 365)/f \rceil$.

3.1.2. Probabilistic Model Inputs.

The first disbursement occurs at the start of the simulation horizon. Each disbursement marks the beginning of a new grant reporting period, so the duration of reporting period k is the time interval between the $(k + 1)$-th and k-th disbursements, defined as the k-th *disbursement interarrival time*, modeled as a random variable and denoted DIT_k.

As discussed in section 2, each grant disbursement is immediately and entirely committed to a procurement order for the product. The quantity purchased then is added to the inventory after a *PLTs* representing the time between order placement and delivery, which is modeled as a random variable denoted PLT_k.

Following the process outlined in section 1.1, the Global Fund assigns a new grant rating to each principal recipient i during each reporting period k, which we denote R^i_k and model as a Markov chain.

Disbursement interarrival times DIT_k, procurement lead times PLT_k and grant ratings R^i_k are the model's uncertain quantities whose effects on inventory and stockout levels are simulated. The following states the input labels used in the remainder of the study for these quantities as well as the exact sections where their estimation procedures are discussed:

$PLT_k:$ Section 3.2.2 $(\text{Input} - \text{PLT} - 4)$

$DIT_k:$ Section 3.2.3 $(\text{Input} - \text{PLT} - 5)$

$R^i_k:$ Section 3.2.4 $(\text{Input} - \text{R} - 6)$

3.1.3. Model Dynamics and Outputs.

Each simulation replication up involves simulation time steps of one day indexed by t and lasts for the time required to satisfy 3 years of demand (nominal duration of

Phase II), which is denoted by T^{ijmlp} in the following. The key model outputs are defined as follows:

Replenishment indicator:

$$O_t = \begin{cases} 1 \text{ if } t = \sum_{\kappa=1}^{k} DIT_\kappa + PLT_k \text{ for some } k; \\ 0 \text{ otherwise} \end{cases}$$

$$(\text{Output} - \text{O} - 7)$$

Inventory evolution:
$$I^{ijmlp}_{t+1} = (I^{ijmlp}_t - 1)^+ + O_t \times \min(f, 3 \times 365 - f$$
$$\times \sum_{\tau=1}^{t-1} O_\tau) \qquad (\text{Output} - \text{I} - 8)$$

Daily lost demand: $S^{ijmlp}_t = (1 - I^{ijmlp}_t)^+$

$$(\text{Output} - \text{S}_t - 9)$$

Total lost demand: $S^{ijmlp} = \sum_{t=1}^{3 \times 365} S^{ijmlp}_t$

$$(\text{Output} - \text{S} - 10)$$

Time to fulfil demand: $T^{ijmlp} = 3 \times 365 + S^{ijmlp}$

$$(\text{Output} - \text{T} - 11)$$

Equation (Output-O-7) defines a daily indicator function associated with the receipt of a procurement order. Equation (Output-I-8) captures inventory dynamics, which are characterized by a normalized demand quantity of 1 unit per day, lost unsatisfied demand, and replenishments occurring on the times defined by (Output-O-7) involving quantities corresponding to the minimum of the nominal disbursement per period f and the total remaining budget (see section 3.1.1). The daily stockout variable S^{ijmlp}_t defined by (Output-S_t-9) provides the cumulative stockout level S^{ijmlp} when summed over 3 years according to (Output-S-10), so that in (Output-T-11) the time T^{ijmlp} required to satisfy 3 years of demand is equal to 3 years plus the stockouts S^{ijmlp} accumulated over that nominal period.

Figure 2 shows a sample simulation replication output for illustration and Table 1 summarizes notation.

3.1.4. Process Modifications Considered.

We use simple modifications of the model defined above to simulate three possible major interventions related to Global Fund financing and procurement processes introduced in section 1.2, as follows:

Instantaneous Replenishment (IR): Immediate delivery of all procurement Orders, for example, from an international or regional buffer stock (warehouse managed by a third-party for the purpose of storing

**Figure 2   Illustrative Simulated Sample Path of Inventory Position
[Color figure can be viewed at wileyonlinelibrary.com]**
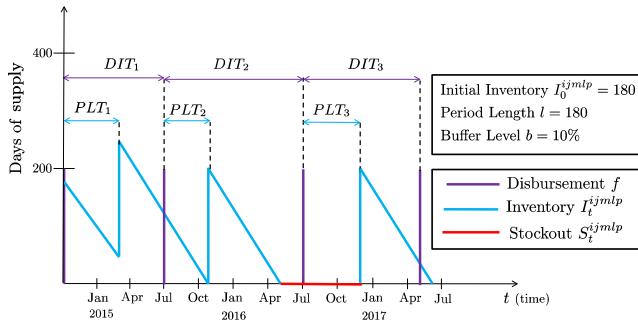
Notes: Illustrative simulated sample path of inventory position I_t^{ijmlp} and stockout level S_t^{ijmlp} over time. Grant monitoring period length l and initial inventory coverage are equal to 180 days. The cash buffer level is $b = 10\%$, so that the per-period disbursement $f = (1 + b)l$ is 198 days. DIT_1, DIT_2, DIT_3 and PLT_1, PLT_2, PLT_3 denote successive realizations of DIT and PLT, respectively. In this illustration, stockout occurs from May to July 2016 due to unavailability of funds and from July to December 2016 due to procurement lead time.

inventory closer to the PR and thus reducing PLTs, Global Fund 2011). This intervention can be captured in the model by replacing the PLT input (Input-PLT-4) with:

$$PLT_k = 0 \qquad (\text{Input} - \text{PLT} - \text{IR} - 12)$$

Bridge Financing (BF): A third party loan for an amount equal to the next anticipated disbursement triggers an advance procurement order placement whenever the DIT exceeds the nominal grant period length l (90 or 180 days), consistent with the

definition of Pledge Guarantee for Health (UNDP 2011). The principal of such loans are then paid back to the third party lender upon grant disbursement. Ignoring financing costs, this intervention can be captured in the model by replacing the inventory replenishment indicator (Output-O-7) with:

$$O_t = \begin{cases} 1 \text{ if } t = \sum_{\kappa=1}^{k-1} DIT_\kappa + \min(DIT_k, l) + PLT_k \\ \text{for some } k; \\ 0 \text{ otherwise} \end{cases}$$

$$(\text{Output} - \text{O} - \text{BF} - 13)$$

Synchronized Financing (SF): In this intervention, nominal grant disbursement amounts are increased to cover one and a half reporting periods, which can be captured in the model by replacing (Input-b-3) with:

$$b = 50\% \qquad (\text{Input} - \text{b} - \text{SF} - 14)$$

To enable meaningful comparisons, overall grant budget is unchanged and disbursements stop when that budget is exhausted. That is, in this intervention the funding schedule is gradually moved forward in time or front-loaded (Natarajan and Swaminathan 2014), but the total amount disbursed over the grant lifecycle remains the same.

3.2. Input Data Estimation
We discuss the datasets used (section 3.2.1), then the estimation procedures for PLTs (section 3.2.2), DITs

Table 1   Model Notation Summary

I/O	Type	Notation	Definition	Value
Inputs	Deterministic	i	Principal recipient	130 distinct recipients
		j	African country	53 distinct countries
		$region(j)$	African region of country j	{North, South, East, West & Central}
		$land(j)$	Indicator if country j is landlocked	{0, 1}
		m	Health product type	(m-1)
		l	Reporting period length (days)	{90, 180}
		p	Grant lifecycle phase	{Phase I, Phase II}
		R_0^i	Initial grant rating	Most frequent for i in dataset
		I_0^{ijmlp}	Initial available inventory	(Input-I-2)
		b	Per-period buffer level	(Input-b-3)
		f	Per-period disbursement amount	$f = (1 + b)l$
		t	Time index	{1, ..., T}
		k	Period index	{1, ..., $\lceil 3 \times 365/f \rceil$}
	Random	PLT_k	Procurement lead time in period k	(Input-PLT-4)
		DIT_k	Disbursement inter-arrival time in period k	(Input-DIT-5)
		R_k^i	Grant rating in period k	(Input-R-6)
Outputs	Random	O_t	Inventory replenishment indicator	(Output-O-7)
		I_t^{ijmlp}	Inventory position at day t	(Output-I-8)
		S_t^{ijmlp}	Lost demand on day t	(Output-S_t-9)
		S^{ijmlp}	Total lost demand over 3 years	(Output-S-10)
		T^{ijmlp}	Time to satisfy 3 years of demand	(Output-T-11)

(section 3.2.3) and grant ratings dynamics (section 3.2.4).

3.2.1. Data Description. The Price and Quality Reporting database is a publicly available database maintained by the Global Fund, where each principal recipient of a procurement grant is required to report all purchases of health products from the following five categories: anti-retroviral drugs, anti-malarial drugs, anti-tuberculosis drugs and prevention of malaria and HIV (Global Fund 2012c). The data contains information about the contents of the procurement order, the order placement and delivery dates and the supplier used (distributor/wholesales or directly from manufacturer). There are 3027 procurement orders delivered from 2002 to the end of 2012.

Our second main data source is a dataset of 2068 disbursements from the Global Fund to principal recipients in 53 countries in Africa from January 2005 to June 2012, obtained from the Global Fund web site (Global Fund 2012d). The variables in this dataset include grant number and disease program funded, disbursement date, reporting period start and end date, and when applicable rating for the previous reporting period data. Disbursements within the first 2 years of each grant lifecycle are identified as Phase I, all others as Phase II.

The third dataset used in the study is the Global Fund's grant data, where each grant approved by the Global Fund is recorded together with the planned start and end dates for both Phase I and Phase II (Global Fund 2012d). There are 461 grants for which Phase I was completed by end of 2012.

3.2.2. Procurement Lead Time Estimation. We obtained historical PLTs from the Price and Quality Reporting dataset and estimated a number of econometric models in order to identify the main factors affecting them (see section A.1 of the online supplement for more details). This analysis led to the selection of a subset of three explanatory variables for PLTs in our model: the product category m; the geographic region of the grant recipient in Africa according to the aggregation of countries used internally by the Global Fund for organizational purposes $region(j) \in \{East, South, North, West \& Central\}$; and whether the receiving country is landlocked $land(j) \in \{0, 1\}$.

The low R^2 in Table A1 suggests a nonlinear relationship between PLTs and these factors. Since we have no prior hypothesis about a particular functional relationship between them, we construct a non-parametric distributional forecast for each combination $(m, region(j), land(j))$ denoted PLT$(m, region(j), land(j))$. The selection of only three explanatory variables was driven by the low minimum number of historical

data points across these combinations (18). As a result, the addition of additional explanatory variables helping to reduce the model's unpredictable variability (e.g., a country's road quality) would have come at a cost to predictive validity.

We validate the predictive accuracy of these forecasts, using repeated out-of-sample evaluation of their predictive accuracy using 1000 randomly selected partitions of the dataset into separate estimation and evaluation sub-samples. For each partition, we use the Kolmogorov-Smirnov test to compare the distributional forecast estimated in-sample and the empirical out-of-sample distribution of PLTs (Arlot and Celisse 2010). As a maximum of 67‰ repetitions (mean 46‰) were rejected at the 5% significance level across all data categories for that test, we conclude that our model of PLTs seems suitably accurate for our purposes, despite admittedly not controlling for any other explanatory variables affecting PLT than the three discussed above.

3.2.3. Disbursement Inter-Arrival Time Estimation. To construct a probabilistic model of DITs for simulation purposes, we followed an approach similar to the one just described for estimating PLTs. Specifically, a regression analysis of historical DITs obtained from the disbursement dataset (see A.2 in the online supplement) along with predictive validity considerations led the selection of the reporting period length $l \in \{90, 180\}$, the country region $region(j)$ and the principal recipient's rating in the previous period R_k^i as the three main explanatory factors for that variable. The simulation of DIT for each principal recipient i in period k thus relies on a distributional forecast constructed for each combination $(l, region(j), R_k^i)$ as a non-parametric estimate of the distribution of DITs over the corresponding subset of historical data.

We likewise evaluated predictive validity for our distributional forecast of DITs using 1000 randomly selected partitions of the dataset into estimation and evaluation sub-samples. As a maximum of 63‰ repetitions (mean 42‰) were rejected at the 5% significance level across all data categories for that test, we conclude that our model of DITs seems suitably accurate for our purposes. We also note that, because of the limited number of data points available for the individual prediction subsets associated with combinations of explanatory variables, this relatively high predictive validity results from the omission of possible additional explanatory variables.

3.2.4. Grant Rating Estimation. Our model simulates successive grant ratings for grants in Phase II as a Markov chain defined for each principal recipient i over the set $\mathcal{R} \triangleq \{\mathcal{A}, \mathcal{B}1, \mathcal{B}2, \mathcal{C}(\mathcal{NR})\}$ (see section 1.1).

We estimated the associated state transition probabilities for each principal recipient from the historical grant rating transitions from the disbursement dataset described in section 3.2.1. Specifically, for every pair of ratings $(r_1, r_2) \in \mathcal{R}^2$ we estimated the transition probability from r_1 to r_2 as the fraction of next periods with rating r_2 when the current period rating for that recipient is r_1 in the dataset. More details and estimation results are provided in section A.3.

3.3. Model Validation

As mentioned in section 1, the present work does not aim methodological contributions but rather empirical contributions relative to the funding process used by the Global Fund and its impact on stockouts of health products in Africa. Consequently, the out-of-sample predictive accuracy of our model with respect to stockout-related indicators constitutes a more appropriate instrument for evaluating its realism than an examination of individual model assumptions. For this reason we now discuss our model's predictive accuracy, and refer the readers to section B.1 in the online supplement for a qualitative discussion of these assumptions.

This validation exercise presents a methodological challenge a priori, because we do not have access to historical data for the total lost demand S^{ijmlp} (see section B.3 for a discussion). Fortunately, however, we are still able to validate our model by using another relevant model output, namely the time T^{ijmlp} required to satisfy the demand associated with the total grant budget determined upfront. Phase II grants are particularly significant here because, unlike Phase I, their planned length is always 3 years without potential for extensions (see section 1.1). As a result we can compare the simulated times T^{ijmlp} with *actual grant implementation lengths*, or total time period over which the funds from a grant were used, which can be estimated from the first and last disbursement dates. This validation measure is meaningful because Phase II grants issued by the Global Fund have a fixed total budget that is determined upfront to precisely cover health program needs for 3 years. As a result, any difference between the actual grant

implementation length and that initial planned period of 3 years indicates a commensurate risk of national stockouts. This can be seen from Equation (Output-T-11), which shows that the difference $T^{ijmlp} - 3 \times 365$ between the simulated grant implementation length and the planned 3-year period provides an estimate for the shortfall S^{ijmlp} in the procurement funds available to cover demand for the health products purchased over the actual grant lifecycle.

We formally define the actual grant implementation length for grant g, $\hat{T}^g$, as the time between the first and last disbursement recorded in the available historical grant records (section 3.1), plus the duration of one grant review period l (90 or 180 days), corrected by a multiplier accounting for the assumed cash buffer level b. Given the information available to us, that definition corresponds to our best estimate of the actual time period over which the funds from that grant were used.

Using the previous definition, we compute the estimated actual implementation lengths for the 429 grants to 62 principal recipients with at least three grants starting before January 1, 2007 recorded in the grant disbursement dataset (out of total of 461 grants, see section 3.2.1). We randomly select around 80% of each principal recipient's grants (347 grants) for estimation of PLT, DIT and rating transition probabilities, and subsequently simulate T^{ijmlp} for each grant in this estimation sample. We perform 5000 replications for each combination of initial inventory I_0 and cash buffer level b. Our baseline simulation parameters (6 months of initial inventory and 10% cash buffer level, see section 3.1) minimize ranked probability score across the 347 in-sample grants (Taylor 2012). These initial conditions and in-sample parameter estimates are then used to simulate T^{ijmlp} for all the out-of-sample grants.

Table 2 below shows average simulated in-sample implementation lengths obtained with these baseline parameters, against average actual in-sample and out-of-sample implementation lengths. While we refer the reader to section 4 for a discussion of the drivers of these implementation lengths, for validation purposes we note here that out-of-sample simulated

Table 2   Simulated and Actual Mean Grant Implementation Lengths

		Observed in-sample			Observed out of sample			Simulated	
	Phase	n	Mean (days)	95% CI (days)	n	Mean (days)	95% CI (days)	Mean (days)	95% CI (days)
90-day Reporting Grants	I	96	841	(822, 860)	25	835	(800, 870)	811	(807, 815)
	II	47	1317	(1285, 1349)	12	1388	(1329, 1447)	1421	(1414, 1428)
180-day Reporting Grants	I	251	779	(771, 787)	57	748	(734, 762)	798	(795, 801)
	II	105	1097	(1095, 1117)	18	1103	(1095, 1148)	1147	(1142, 1152)

Notes: Mean implementation length and 95% CI by phase and reporting frequency for actual in-sample and out-of-sample actual data against simulated predictions generated from in-sample data using baseline parameter values. Variable n represents sample size.

implementation lengths of 90-day grants (resp. 180-day grants) are on average only 2.9% shorter (resp. 2.4% longer) than actual out-of-sample values in Phase I (resp. Phase II). In addition, for 180-day reporting grants these average relative prediction errors are only +6.7% in Phase I and +4.0% in Phase II.

The online supplement also provides a comparison of the distribution of simulated times T^{ijmlp} with the distribution of actual implementation lengths $\hat{T}^g$ of the 82 out-of-sample grants. This analysis leads to the conclusion that we cannot reject, at the 10% significance level, the hypothesis that actual out-of-sample observations of the procurement funds missing to cover demand for health product over a grant lifecycle follow the simulated distribution of the same quantity (see section B.3).

These results suggest that despite a number of simplifying assumptions, the simulation model and associated data estimation procedures defined in sections 3.1 and 3.2 satisfactorily capture the stockout risks associated with the Global Fund funding and procurement processes for the purpose of this study.

4. Results and Discussion

The estimation results for the distributional forecasts of the probabilistic input variables DITs and PLTs reported in sections A.2 and A.1 of the online supplement show that both financial and physical flows related to the supply continuity of health products purchased with Global Fund grants exhibit substantial unpredictable variability (average coefficient of variation of 0.657 and 0.508 for DITs and PLTs, respectively). Furthermore, a high proportion of historical observations have DIT longer than the grant reporting period, particularly for 90-day period grants. This raises concerns that the Global Fund's disbursement schedules may lack reliability and are slower than the health programs they are designed to support. To investigate these issues and quantify their impact on stockouts, in the following section we

discuss the results of extensive simulation experiments performed with the model described in the previous section, and their implications on the motivating questions mentioned in section 1. Specifically, we examine the impact of grant reporting frequency and geographic region of recipients in sections 4.1 and 4.2, respectively, then evaluate potential interventions in section 4.3. Within each subsection, we first present the relevant empirical results and then discuss their implications.

4.1. Effect of Reporting Frequency on Stockout Risks

A first set of experiments assumed baseline parameters and grants in Phase II. They involved simulation runs for every (i, j, l, m) combination of principal recipient, country, reporting frequency and product type in our dataset, or 541 data instances.[2] Table 3 includes the resulting estimates of the average proportion of demand lost over 3 years, aggregated over each African region, where the aggregations across principal recipients in the same country and across countries in the same African region, were performed with weights equal to the corresponding relative volumes of funding disbursed by the Global Fund. We also refer the reader to Table C1 in the online supplement for more detailed results at the country level.

A first observation from Table 3 is the high absolute level of predicted national stockout risks for 90-day grants, with an average simulated proportion of demand lost over 3 years of 28.7% across 90-day grants, reaching a maximum of 49.3% for 90-day tuberculosis grants in East Africa. These high predicted stockout risks are consistent with the independent field observations of widespread stockouts in countries receiving Global Fund financing, as reported in section 1.

The results shown in Tables 3 and C1 also indicate that the grants for which concerns about performance or implementation risks led the Global Fund to use a shorter reporting period of 90 days as opposed to

Table 3 Simulated Average Proportion of Demand (%) Lost over 3 Years for Baseline Scenario

| | | Health product type m | | | | | | | | |
| | | Anti-malarial | | Anti-retroviral | | Anti-tuberculosis | | HIV prevention | | Malaria prevention | |
	Reporting frequency (days) l	90	180	90	180	90	180	90	180	90	180
African region	North	29.32	3.07	29.83	2.94	31.55	4.61	29.94	2.22	31.58	4.51
	East	44.49	11.27	48.11	11.17	49.31	11.75	47.21	10.13	48.53	10.32
	South	32.14	5.02	29.31	4.24	38.27	6.34	28.03	3.65	30.84	4.27
	West & Central	21.43	2.60	21.64	2.03	22.64	2.25	20.86	1.70	21.69	2.47

Notes: Simulated average proportion of demand (%) lost over 3 years for baseline scenario in Phase II. Results based on 5000 replications ensuring the length of the 95% confidence interval is <1% of the estimated expected stockouts for each parameter combination. Results for different grants within each African region aggregated using weights proportional to total grant amounts.

180 days clearly faced substantially higher national stockout risks: with an average 28.7% lost demand for 90-day vs. 5.3% for 180-day reporting grants, expected lost demand for 90-day reporting grants ranges from 4 to 20 times that of 180-day grants for the same disease and country. The result that shorter grant reporting periods are associated with greater stockout risks may not seem surprising *per se*, because shorter reporting periods can be seen as imposing more stringent constraints on cash availability. However, what is both surprising and important here from a practical standpoint is the substantial extent to which 90-day grants impact stockout risks relative to 180-day grants. In other words, over the first 11 years of the Global Fund there was a clear effective trade-off between the extent of its performance monitoring activities and the effectiveness of the funds it disbursed.

It is legitimate to ask whether the difference in expected stockouts between 90-day and 180-day grants reported in Table 3 may be due to unobservable intrinsic risk factors that could have influenced the decision to use 90-day or 180-day mechanisms for these grants, rather than the reporting mechanism itself. While performing a controlled experiment was not a feasible option in this setting, some observations support the hypothesis of a causal impact of grant monitoring frequency on stockout risks—see section C.1 in the online supplement.

These results have implications for policy and practice. The substantially higher stockout risks associated with 90-day grants warrant a detailed examination of whether the relative benefits of these grants in terms of management incentives are commensurate. It is noteworthy that, because of the Global Fund's historical practices, the long DITs estimated in our study may have been caused by issues affecting any of the activities associated with the execution of a grant, including activities having nothing to do with the procurement of health commodities. For example, a delay with the complete documentation of expenses linked to the construction of a health clinic or an advertising program on condoms could conceivably postpone an incoming disbursement to be used primarily for procuring medicines.

Because the short-term public health impact of delays affecting medicine procurement may be quite different from that of delays affecting other grant components, it would seem beneficial for the Global Fund to manage the schedule of procurement-related disbursements in a specific manner. We note that other donors also implementing performance-based funding principles such as the World Bank's Health Results Innovation Trust Fund and the GAVI Alliance already separate payments into fixed/predictable and performance-based portions, presumably for the

same reasons (Fan et al. 2013). Applying this model to funding for procurement, the Global Fund could further protect procurement-related disbursements by reducing their dependence on performance considerations, particularly when these considerations are unrelated to procurement.

More generally, the trade-off between fund effectiveness and financing predictability on the one hand and performance incentives on the other hand could be systematically managed in a segmented manner across different grant components, increasing overall efficiency. Alternative mechanisms for preserving patient access to medicines without compromising fund integrity include letters of credit directly issued to manufacturers and imposing the use of central procurement services similar to those currently known as Pooled Procurement Mechanism. It is not clear that the Global Fund systematically uses such alternative mechanisms when performance concerns related to procurement arise. Interventions including vendor-managed inventory are conceivable, but may be challenging to implement in this context because the level of trust between buyers and suppliers seems to strongly influence the success of such relationships (Claassen et al. 2008), and such trust may be difficult to establish in the presence of creditworthiness concerns.

4.2. Effect of Geographic Location on Stockout Risks

Another important observation from Table 3 is that predicted stockout risks are strongly correlated with the geographic region of receiving countries. Specifically, expected lost demand for 90-day (resp. 180-day) reporting grants range from around 21% (resp. 2%) in West & Central Africa to around 49% (resp. 11%) in East Africa. In addition, these predicted stockout risks are quite consistent for each region across product types, suggesting that the geographic region is a more important driver of stockouts than the type of product being purchased (some comments on the impact of product type are still included in section C.2 of the online supplement). Finally, an examination of the more detailed country-level results provided in section C of the online supplement reveals that the variability of predicted stockout risks across countries within the same geographic region is quite limited.

These results seem hard to rationalize from a public health or performance monitoring standpoint. It is possible that these geographic regions should actually coincide with some intrinsic features of recipient countries that would similarly affect the processes used for grant performance evaluations and disbursements. Given the heterogeneity of countries within these regions along many dimensions however

(Berenguer et al. 2016), this explanation does not seem plausible. Rather, we have used for this analysis the exact definition of geographic regions used by the Global Fund for reporting purposes, and these regions are also reflected in its internal organizational structure—for example, fund portfolio managers are almost always responsible for countries within a single region. Thus, these observations strongly suggest that the predicted stockouts are primarily driven by organizational features and specific processes used by the Global Fund (which are common to countries in the same region but differ across regions), as opposed to underlying risk factors associated with individual countries (which presumably differ widely across countries in the same region).

This explanation aligns with observations made independently by Fan et al. (2013) on the basis of both econometric analysis of historical grant scores and detailed case studies of Global Fund decisions for several specific countries. Specifically, these authors highlight the lack of transparency and apparent subjectivity affecting the relationship between grant evaluation scores and actual disbursement decisions made by Global Fund teams.

Based on this collective evidence, it seems important for the Global Fund to develop processes and guidelines for evaluating and acting upon grant management performance that are more objective, globally scalable, and immune from organizational idiosyncrasies. Beyond the Global Fund, these observations also seem relevant to any other global health funding organization implementing or considering a performance-based funding model. Specifically, they highlight an important tension associated with a decentralized organization structured around geography when implementing a performance-based funding model, namely the benefits of in-depth local knowledge by teams vs. the challenges of implementing a uniform and objective set of evaluation criterias across teams.

4.3. Potential Impact of Policy Interventions

While it is expected that the interventions considered in this study (reducing PLTs, bridge financing and increasing cash buffers) should all result in some reduction of stockout risks, the goal of our study is to evaluate the relative benefits of these different interventions. To that end, we conducted a second larger set of numerical experiments where the simulation runs were not only defined by the combination (i, j, l, m) of principal recipient, country, reporting frequency and product type as in the first set of experiments (see section 4.1), but also which one of the three potential interventions discussed in sections 1.2 and 3.1.4 is being considered. While the primary performance metric estimated was the proportion of demand lost over 3 years as before, we also conducted additional sensitivity analysis experiments to investigate the impact of the evaluation period duration (see below). Table 4 reports summary statistics related to these experiments, where the results of individual simulation runs have been aggregated across countries in the same geographic region (using the aggregation method described in section 4.1) and across product types (see methodological note in Table 4).

As seen in section 3.1.4 of the model definition, the instantaneous replenishment intervention effectively amounts to eliminating PLTs from the inventory dynamics (i.e., setting the PLT_k distributions to zero) while leaving the financial flows (i.e., the DIT_k distributions) unchanged. This is a simplified model of an intervention that would consist in practice of setting up regional buffer stocks that principal recipients may access on a short notice. The simulated results for this intervention are also meaningful because

Table 4   Simulated Average Proportion of Demand Lost (%) over 3 Years for Baseline and Policy Interventions

| | | Legacy practice | | | | Intervention | | | | | |
| | | Phase I | | Phase II | | Instantaneous replenishment | | Bridge financing | | Synchronized financing | |
	Reporting frequency (days) *l*	90	180	90	180	90	180	90	180	90	180
African region	North	29.9	1.1	29.7	3.4	22.7	0.8	26.4	2.1	13.7	1.06
	East	55.1	5.8	40.8	9.1	33.5	3.3	37.5	6.8	27.7	3.08
	South	29.3	2.3	24.9	3.3	20.1	1.1	22.3	2.0	10.1	0.86
	West & Central	23.7	2.0	19.5	2.4	13.5	0.5	16.0	1.7	6.5	0.85

Notes: Simulated average proportion of demand lost (%) over 3 years for baseline scenario and three potential interventions. Results based on 5000 replications ensuring the length of the 95% confidence interval is <1% of the estimated expected stockouts for each parameter combination. Results across different drugs for each principal recipient are aggregated using weights proportional to the following: the number of reported malaria cases in 2010 for anti-malaria and malaria prevention drugs; the sum of people in need of and on ARV treatment for ARV drugs; the number of people living with AIDS for HIV prevention drugs; and the number of new TB cases in 2011 for anti-TB drugs (The Henry J. Kaiser Family Foundation 2012). Obtained results for different principal recipients within each African region are then aggregated using weights proportional to the sum of grant amounts.

comparing them with the baseline results provides an estimation of the specific impact of delays due to PLTs as opposed to DITs. Indeed, the results shown in Table 4 suggest that instantaneous replenishment would decrease expected stockouts between 4.8 and 7.3 percentage points (pp) for 90-day grants, and between 1.9 and 5.8pp for 180-day grants. While this intervention would thus reduce stockouts to minimal levels for 180-day grants (under 3.5% of demand over 3 years), it would unfortunately leave 90-day grants with high levels of stockout exposure (between 13.5% and 33.5% of demand over 3 years). These results can be explained by the greater discrepancy observed for 90-day grants between disbursement amounts (designed to cover demand for the nominal grant monitoring period) and the actual time between consecutive disbursements (see section A.2 in the online supplement). Consequently, the estimated stockouts for 90-day grants are primarily driven by DITs as opposed to PLTs, which explains the poor targeting efficiency of the instantaneous replenishment intervention.

The results shown in Table 4 also suggest that the impact of bridge financing would be limited. Specifically, bridge financing achieves even lower stockout reductions than instantaneous replenishment, also leaving stockout risks for 90-day grants at relatively high levels (minimum 16% in West & Central Africa). The explanation is that bridge financing, while making some funds available earlier, changes neither the amount nor the frequency of the underlying Global Fund disbursement schedule. That is, bridge funds provide coverage for a nominal grant reporting period (i.e., 90 or 180 days), but because their disbursement remains linked to the schedule of actual Global Fund disbursements which are separated by the actual DITs, the time difference between nominal grant review period and actual DITs remains ultimately unfunded. As a result, the bridge financing policy considered here, which is consistent with the Pledge Guarantee for Health scheme promoted by the United Nations Development Fund (see section 3.1 and UNDP 2011), does not address the structural problem of disbursement timing and amount and does not constitute a reliable process for preventing stockouts. Furthermore, an actual implementation of bridge financing would likely entail additional interest and financing costs which we conservatively ignore here. Alternative mechanisms are conceivable, but any implementation increasing disbursement frequency would expose third-party lenders to financial liability and risks accumulating over time.

Finally, synchronized financing is the only considered intervention substantially reducing stockout risks for 90-day grants—by between 13pp in West &

Central Africa and 16pp in North Africa. In addition, its associated stockout risks are also low for 180-day grants, and comparable to those achieved by instantaneous replenishment. The explanation of this relatively high potential impact is that synchronized financing directly addresses the core issue that planned disbursement amounts designed to cover demand for review periods of fixed duration (e.g., 90 or 180 days) were not adjusted to reflect longer actual time periods between disbursements (see discussion of the historical Global Fund disbursement process in section 1.1). Indeed, the additional cash buffers associated with synchronized financing correct this by effectively making disbursement amounts commensurate with empirical DITs. Thus, the variability of these empirical DITs across regions thus explains why the uniform additional cash buffer level of 50% assumed for the synchronized financing policy in the experiments reported in Table 4 has an impact which varies across regions.

To further investigate this issue, additional experiments on the synchronized financing policy reported in Figure 3 show that simulated stockouts are sensitive to the cash buffer level. Specifically, increasing cash buffer levels generally has a substantial marginal impact on stockouts until levels of approximately 70% for 90-day grants and 20% for 180-day grants. These thresholds correspond to a probability of approximately 0.4 for the event that $DIT_k > f$: beyond that point the occurence of stockouts in a given period rapidly become less likely given the DIT_k distribution tail, so the marginal benefits of a cash buffer reduce substantially. This sensitivity analysis may inform the choice of a specific cash buffer level, achieving a good balance between stockout risks and the financial exposure and/or changes of incentives associated with disbursing more funds upfront. On the basis of these results, one could conceivably also consider a potential implementation of synchronized financing involving different cash buffer levels across regions. However, we stress that these results are mostly driven by the underlying differences in DITs across regions resulting from different practices across teams within the Global Fund (see section 4.2). As a result, addressing these organizational differences directly would seemingly constitute a more durable solution than accomodating them through segmented cash buffer levels.

Finally, additional experiments reported in Figure 4 suggest that the stockout reductions associated with these three interventions is sensitive to the time horizon considered. After an initial period of 6 months corresponding to the initial inventory assumed, the fraction of lost demand generally increases over time for both reporting periods and all interventions, with the exception of synchronized financing for 180-day

Figure 3   Fraction of Lost Demand for Different Cash Buffer Levels [Color figure can be viewed at wileyonlinelibrary.com]
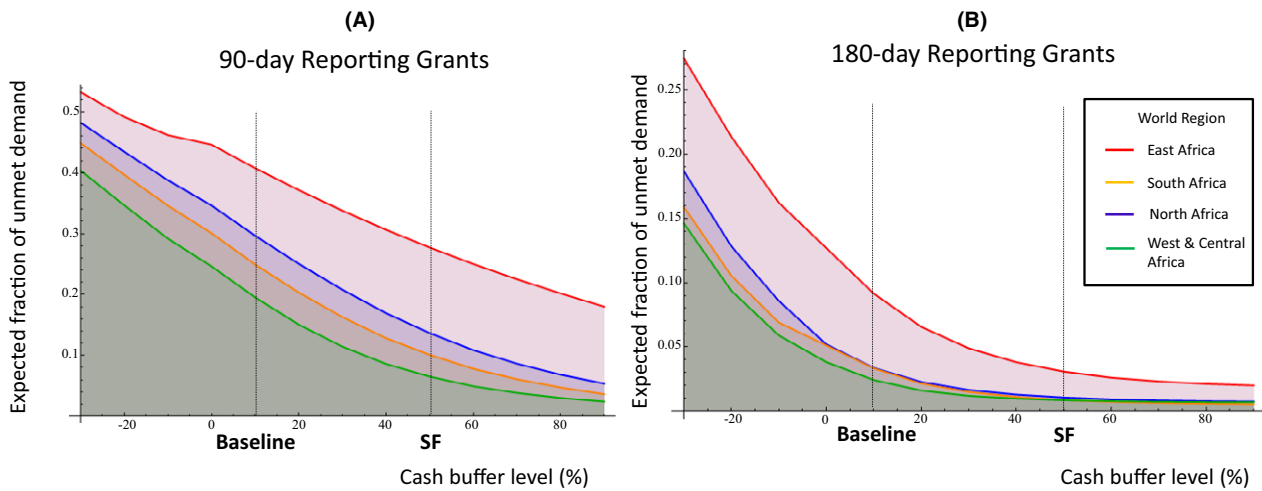
Notes: Average simulated fraction of lost demand for different cash buffer levels in each African region: 90-day reporting grants (A) and 180-day reporting grants (B). Methodology for aggregating results is identical to that described in the notes of Table 4. Baseline parameters in Phase II are assumed except for cash buffer level. Highlighted cash buffer levels of 10% and 50%, respectively, correspond to the baseline scenario and synchronized financing (SF) intervention.

Figure 4   Fraction of Lost Demand over Various Time Horizons [Color figure can be viewed at wileyonlinelibrary.com]
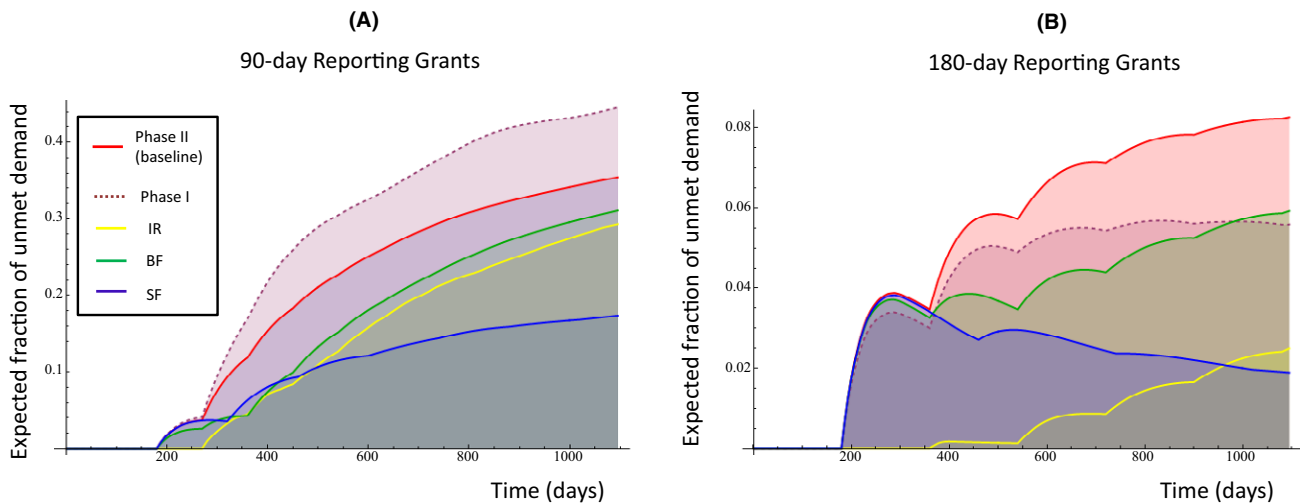
Notes: Average simulated fraction of lost demand over various time horizons under baseline parameters and different interventions, aggregated across African countries: 90-day reporting grants (A) and 180-day reporting grants (B). Methodology for aggregating results is identical to that described in the notes of Table 4. Initial inventory of 6 months and cash buffer level of 10% assumed for all scenarios except synchronized financing where the cash buffer level is 50%. Scenarios are Phase II (baseline), Phase I, IR (instantaneous replenishment), BF (bridge financing) and SF (synchronized financing).

grants. This reflects the fact that, in all cases but the exception noted, provided funds are insufficient to cover the average time between consecutive disbursement (i.e., $\mathbb{E}[DIT_k] > f$), so that stockouts accumulate over time.

Contrasting with the 3-year results shown in Table 4, expected lost demand under instantaneous replenishment is lower than with synchronized financing up to day 475 (resp. 960) for 90-day (resp. 180-day) reporting grants. This is because instantaneous replenishment advances the delivery of the first replenishment relative to synchronized financing, which substantially reduces stockouts at the beginning of the time horizon. Instantaneous replenishment could thus become a sensible intervention for grants with shorter durations than the current ones,

however its implementation seems more involved than synchronized financing and so associated costs should be carefully examined.

A key implication of these results for the Global Fund is that adjusting disbursement amounts through a more systematic use of cash buffers reflecting actual disbursement schedules would substantially benefit public health, and appears more effective in the long term than the other possible interventions considered here. Table 4 suggests countries and regions to be targeted in priority for this intervention, and the high absolute level of estimated stockouts suggest that this may be time-sensitive in many cases.

5. Conclusion

Our findings provide new evidence on the relationship between global health initiatives and national health systems by identifying and characterizing the link between the Global Fund's financing and disbursement processes and national drug stockout risks over the past decade (WHO 2009). These results complement observational studies on health product stockouts in Africa (e.g., Oliynyk 2011, Pasquet et al. 2010) and qualitative studies discussing their causes by providing a validated model generating quantitative predictions of stockout risks and characterizing the role of disbursement and procurement variability (Shretta and Yadav 2012). Finally, our study provides hitherto unavailable quantitative predictions of the impact of potential interventions for reducing these risks, and point to several process redesign opportunities. Beyond the Global Fund, these findings also seem useful to other global health organizations interested to identify and leverage learnings from the first large-scale implementation of the performance-based funding model.

Specifically, we find that the higher grant reporting frequency is a substantial driver of stockout risks, so that over the first 11 years of the Global Fund there was a clear effective trade-off between the extent of its performance monitoring activities and the effectiveness of the procurement funds it disbursed. This more generally shows the importance of properly accounting for the resource and time requirements of performance monitoring activities when planning the execution of performance-based funding grants. Our results show that East Africa faced much higher and West & Central Africa much lower stockout risks than the rest. This suggests that the African regions used internally by the Global Fund for organizational purposes likely constitute a substantial driver of stockout risks variability because of idiosyncratic and region-specific differences in evaluation and/or disbursement decision processes. This highlights an organizational tension that seems important to

carefully manage as part of the performance-based funding model, namely the benefits of in-depth local knowledge by teams dedicated to limited geographic areas vs. the challenges of implementing a uniform and objective set of evaluation criterias across teams. Finally, we find that adjusting disbursement amounts, using cash buffers commensurate with the actual duration of monitoring periods has the potential to reduce expected stockouts more significantly than regional buffer stocks and bridge financing (see section 4.3).

The substantial stockout risks imposed upon many Global Fund recipients since 2002 that are highlighted in this study seem significant to global health. This may motivate a more extensive redesign of the Global Fund's funding model than the changes we could evaluate in this study on the basis of historical DIT data, and may have motivated the new funding model initiative announced by the Global Fund in 2013 (Global Fund 2013). Indeed, all the interventions considered in our quantitative study essentially assume that the process used by the Global Fund for the purpose of monitoring the performance of grant recipients would remain unchanged relative to the collection period of our DIT data from 2002 to 2013, or at least that any changes considered would not impact the distribution of DITs. However, our study and model do provide a framework for thinking about further redesign opportunities for this process in a systematic manner.

For example, the assumed lack of correlation between inventory level and DIT points to the current lack of centralized country stock level information accessible to the Global Fund on a routine basis for the health products that it is funding. This situation is particularly problematic when Global Fund managers are confronted with several competing solicitations for expediting disbursements or allocating limited stock or funds available to them in the short term, as this lack of information may contribute to inefficient decisions with severe consequences. This information scarcity also complicates the development of proactive and forward-looking approaches for allocating funds and resources, contrasting with the reactive "fire-fighting" environment generated by emergency solicitations of recipient countries facing an existing or imminent stock-out crisis (a key motivation for the support of this study by the Global Fund). Finally, this lack of reliable centralized stock level information makes it difficult for the Global Fund to evaluate the performance of recipient countries in relation to inventory management, and therefore hampers its core performance monitoring function. More generally, this suggests that other global health organizations involved in the allocation of procurement funds as part of the

performance-based funding model may find it particularly beneficial to develop some visibility of central inventory levels across recipients.

We see several future research opportunities related to this work. One would be a deeper study of the longer-term evolution of grant reporting frequency and ratings across multiple successive grants received by the same principal recipient, possibly using a system dynamics model. Efficiency analysis could also shed more light on the specific variables affecting PLTs and DITs. The process by which incoming fund disbursements are split between procurement orders for different products seems worthy of study, as is the possible coordination between multiple funding streams. Finally, a microeconomic model could generate useful knowledge on the relationship and possible contractual forms between a donor and a recipients in the context of performance-based financing.

Notes

[1]Some analysis of the historical drivers of grant reporting frequency is discussed in section D.2 in the online supplement.

[2]The number of data instances is larger than the number of grants in the dataset since HIV and malaria grants can be used for the purchase of both treatment and prevention drugs.

References

ALMA. 2011. African Leaders Malaria Alliance, evidence to policy initiative, clinton health access initiative. Maintaining the gains in malaria control. Country brief: Rwanda, September 2011.

Arlot, S., A. Celisse. 2010. A survey of cross-validation procedures for model selection. *Stat. Surveys* **4**: 40–79.

Atun, R., F. M. Knaul, Y. Akachi, J. Frenk. 2012. Innovative financing for health: What is truly innovative? *Lancet* **380**: 2044–2049.

Berenguer G., A. Iyer, P. Yadav. 2016. Disentangling the efficiency drivers in country-level global health programs: An empirical study. *J. Oper. Manag.* **45**: 30–43.

Brugha, R., M. Donoghue, M. Starling, P. Ndubani, F. Ssengooba, B. Fernandes, G. Walt. 2004. The Global Fund: Managing great expectations. *Lancet* **364**: 95–100.

Buzacott, J. A., R. Q. Zhang. 2004. Inventory management with asset-based financing. *Management Sci.* **50**(9): 1274–1292.

Celasun, O., J. Walliser. 2007. Predictability of aid: do fickle donors undermine economic development? 46th Panel Meeting of Economic Policy, Lisbon 2007.

Center for Global Development. 2013. Report: More health for the money. Available at http://international.cgdev.org/sites/default/files/More-Health-for-the-Money.pdf (accessed date February 12, 2014).

Chao, X., J. Chen, S. Wang. 2008. Dynamic inventory management with cash flow constraints. *Nav. Res. Log.* **55**(8): 758–7768.

Claassen, M., A. Van Weele, E. Van Raaij. 2008. Performance outcomes and success factors of vendor managed inventory (VMI). *Supply Chain Manag. Int. J.* **13**: 406–414.

Cohen, J., I. Singh, M. O'Brien. 2008. Predicting Global Fund grant disbursements for procurement of artemisin-based combination therapies. *Malaria J.* **7**: 200.

Fan, Y., D. Duran, R. Silverman, A. Glassman. 2013. Performance-based financing at the Global Fund to fight AIDS, Tuberculosis and Malaria: An analysis of funding practice, 2002. *Lancet Global Health* **1**: e161–e168.

Foster, S., R. Laing, B. Melgaard, M. Zaffran. 2006. Ensuring supplies of appropriate drugs and vaccines. D. Jamison, J. Breman, A. Measham, G. Alleyne, M. Claeson, D. Evans, P. Jha, A. Mills, P. Musgrove, eds. *Disease Control Priorities in Developing Countries*, 2nd edn. Oxford University Press, New York, 1323–1338.

Gallien, J., N. Leung, P. Yadav. 2016. Public distribution of essential medicines in Zambia: A heuristic, dataset, and validated simulation model. Working paper.

Garnett, G., S. Cousens, T. Hallett, R. Steketee, N. Walker. 2011. Mathematical models in the evaluation of health programmes. *Lancet* **378**: 515–525.

Gong, X., X. Chao, D. Simchi-Levi. 2014. Dynamic inventory control with limited capital and short-term financing. *Nav. Res. Log.* **61**(3): 184–201.

Hawkes, N. 2011. Gains in malaria control will be lost if funding stalls. *Br. Med. J.* **343**: d5970.

IHME (Institute for Health Metrics and Evaluation). 2014. *Financing Global Health 2013: Transition in an Age of Austerity.* IHME, Seattle, WA.

Kouvelis, P., J. Li. 2008. Flexible backup supply and the management of lead-time uncertainty. *Prod. Oper. Manag.* **17**: 184–199.

Kraiselburd, S., P. Yadav. 2011. Supply chains and global health: An imperative for bringing operations management scholarship into action. *Prod. Oper. Manag.* **22**(2): 377–381.

Lane, C., A. Glassman. 2008. Smooth and predictable aid for health: A role for innovative financing? *Brookings Institution* Global Health Financing Initiative working paper.

Levi, R., G. Perakis, G. Romero. 2016. On the effectiveness of uniform subsidies in increasing market consumption. *Management Sci.* **63**(1): 40–57.

Levine, R., J. Pickett, N. Sekhri, P. Yadav. 2008. Demand forecasting for essential medical technologies. *Am. J. Law Med.* **34**: 225–255.

Lu, C., C. M. Michaud, K. Khan, C. Murray. 2006. Absorptive capacity and disbursements by the Global Fund to Fight AIDS, Tuberculosis and Malaria: Analysis of grant implementation. *Lancet* **368**: 483–488.

Ministry of Health, Uganda. 2012. Stock Status Report as at 1st November 2012. Available at www.health.go.ug/docs/ssr_nov_12.pdf (accessed date January 5, 2013).

Natarajan, K. V., J. M. Swaminathan. 2014. Inventory management in humanitarian operations: Impact of amount, schedule, and uncertainty in funding. *Manuf. Serv. Oper. Manag.* **16**: 595–603.

Oliynyk, I. 2011. Stock-outs in Global Fund's supported programs. Presentation, Workshop on Procurement and Supply Management of HIV and TB Products, Nairobi, Kenya.

Pasquet, A., E. Messou, D. Gabillard, A. Minga, A. Depoulosky, S. Deuffic-Burban, E. Losina, K. A. Freedberg, C. Danel, X. Anglaret, Y. Yazdanpanah. 2010. Impact of drug stock-outs on death and retention to care among HIV-infected patients on combination antiretroviral therapy in Abidjan, Cte d'Ivoire. *PLoS ONE* **5**: e13414.

Radalet, S., B. Siddiqi. 2007. Global Fund grant programmes: An analysis of evaluation scores. *Lancet* **369**: 1807–1813.

Shretta, R., P. Yadav. 2012. Stabilizing supply of artemisinin and artemisinin-based combination therapies in an era of widespread scale-up. *Malaria J.* **11**: 399.

Song, J. S. 1994. The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Sci.* **40**(5): 603–613.

Song, J. S., H. Zhang, Y. Hou, M. Wang. 2010. The effect of lead time and demand uncertainties in (*r,q*) inventory systems. *Oper. Res.* **58**(1): 68–80.

Taylor, J. 2012. Density forecasting of intraday call center arrivals using models based on exponential smoothing. *Management Sci.* **58**(3): 534–549.

Taylor, T. A., W. Xiao. 2014. Subsidizing the distribution channel: Donor funding to improve the availability of malaria drugs. *Management Sci.* **60**(10): 2461–2477.

The Global Fund. 2006. *Quantification from a TB Perspective, Global Drug Facility*. Technical Briefing Seminar, Copenhagen.

The Global Fund. 2009. Stock out episodes in Global Fund grants. Survey among Fund Portfolio Managers, January 2008–December 2009.

The Global Fund. 2011. Report of the Portfolio and Implementation Committee, 25th Board meeting, Ghana.

The Global Fund. 2012a. Who we are. Available at http://www.theglobalfund.org/en/about/whoweare/ (accessed date May 24, 2012).

The Global Fund. 2012b. Guidelines for completing the PR "ongoing progress update and disbursement request", and LFA"ongoing progress review and disbursement recommendation".

The Global Fund. 2012c. Price and Quality Reporting Database. Available at http://pqr.theglobalfund.org (accessed date October 15, 2012).

The Global Fund. 2012d. Disbursement in detail (raw data) and Grants in detail (raw data). Available at http://portfolio.theglobalfund.org/en/Home/Index (accessed date June 27, 2012).

The Global Fund. 2013. News Flash, Issue 16. Transition to annual grant disbursement. March 21, 2013. Available at http://www.theglobalfund.org/en/blog/31717/ (accessed date March 23, 2013).

The Henry J. Kaiser Family Foundation. 2012. U.S. Global Health Policy. Available at http://www.globalhealthfacts.org/data/topic/map.aspx?ind=30&by=Location (accessed date November 28, 2012).

The PLoS Medicine Editors. 2009. Time for a "Third Wave" of Malaria Activism to tackle the drug stock-out crisis. *PLoS Med.* **6**(11): e1000188.

Tougher, S., Y. Ye, J. Amuasi, I. A. Kourgueni, R. Thomson, C. Goodman, A. G. Mann, R. Ren, B. A. Willey, C. A. Adegoke, A. Amin, D. Ansong, K. Bruxvoort, D. A. Diallo, G. Diap, C. Festo, B. Johanes, E. Juma, A. Kalolella, O. Malam, B. Mberu, S. Ndiaye, S. B. Nguah, M. Seydou, M. Taylor, S. T. Rueda, M. Wamukoya, F. Arnold, K. Hanson. 2012. Effect of the Affordable Medicines Facility–malaria (AMFm) on the availability, price, and market share of quality-assured artemisin-based combination therapies in seven countries: a before-and-after analysis of outlet survey data. *Lancet* **380**: 1916–1926.

UNDP. 2011. United Nations Development Fund. Pledge Guarantee for Health. Available at http://www.unfoundation.org/what-we-do/campaigns-and-initiatives/pledge-guarantee-for-health/our-work/ (accessed date January 4, 2013).

United Nations. 2013. The Millennium Developement Goals Report. Available at http://www.un.org/millenniumgoals/pdf/report-2013/mdg-report-2013-english.pdf (accessed date February 12, 2014).

Wang, Y., B. Tomlin. 2009. To wait or not to wait: Optimal ordering under lead time uncertainty and forecast updating. *Nav. Res. Log.* **56**: 766–779.

WHO. 2004. *The World Medicines Situation Report*. WHO, Geneva.

WHO. 2009. World Health Organization Maximizing Positive Synergies Collaborative Group. An assessment of interactions between global health initiatives and country health systems. *Lancet* **373**: 2137–2169.

WHO. 2013a. *Global Tuberculosis Report*. WHO, Geneva, Switzerland.

WHO. 2013b. *World Malaria Report*. WHO, Geneva, Switzerland.

Yadav, P. 2007. Analysis of the public, private and mission sector supply chains for essential drugs in Zambia. Working paper.

Yu, D., Y. Souteyrand, M. Banda, J. Kaufman, J. Perriens. 2008. Investment in HIV/AIDS programs: Does it help strengthen health systems in developing countries? *Global. Health* **4**(8). doi:10.1186/1744-8603-4-8.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix A: Model Input Data.
Appendix B: Model Assumptions.
Appendix C: Additional Model Results.
Appendix D: Additional Data Analysis.