

Global Fund grant programmes: an analysis of evaluation scores



Steven Radelet, Bilal Siddiqi

Summary

Background The Global Fund to Fight AIDS, Tuberculosis and Malaria evaluates programme performance after 2 years to help decide whether to continue funding. We aimed to identify the correlation between programme evaluation scores and characteristics of the programme, the health sector, and the recipient country.

Methods We obtained data on the first 140 Global Fund grants evaluated in 2006, and analysed 134 of these. We used an ordered probit multivariate analysis to link evaluation scores to different characteristics, allowing us to record the association between changes in those characteristics and the probability of a programme receiving a particular evaluation score.

Findings Programmes that had government agencies as principal recipients, had a large amount of funding, were focused on malaria, had weak initial proposals, or were evaluated by the accounting firm KPMG, scored lowest. Countries with a high number of doctors per head, high measles immunisation rates, few health-sector donors, and high disease-prevalence rates had higher evaluation scores. Poor countries, those with small government budget deficits, and those that have or have had socialist governments also received higher scores.

Interpretation Our results show associations, not causality, and they focus on evaluation scores rather than actual performance of the programmes. Yet they provide some early indications of characteristics that can help the Global Fund identify and monitor programmes that might be at risk. The results should not be used to influence the distribution of funding, but rather to allocate resources for oversight and risk management.

Introduction

The Global Fund to Fight AIDS, Tuberculosis and Malaria has become one of the world's largest funders of programmes fighting these three diseases. Just 5 years after its founding, it contributes two-thirds of all donor resources to fight tuberculosis, half for malaria, and one-quarter for HIV/AIDS.¹ By late 2006, the organisation had approved proposals worth \$6·8 billion for 448 programmes in 136 countries, and disbursed over \$3 billion.²

The organisation's grant approval process is unusual. Programmes are initiated by Country Coordinating Mechanisms (CCMs)—partnerships of governments, non-governmental organisations (NGOs), faith-based organisations, and civil society groups in each country—which submit proposals to the Global Fund Secretariat. Each proposal is reviewed by a technical review panel—an independent group of experts that makes recommendations to the Global Fund Board for funding decisions. Approval is far from automatic: the board approves less than 50% of submitted proposals.

After a proposal is approved, the secretariat negotiates a grant agreement with a principal recipient—an organisation nominated by the CCM to receive funds, implement programmes, and disburse funds to sub-recipients. The most common principal recipients are ministries of health, national AIDS (or tuberculosis or malaria) coordinating bodies, church groups, NGOs, or the United Nations Development Programme (UNDP). To monitor programmes, the secretariat (which has no

staff in recipient countries) hires a Local Fund Agent (LFA)—typically a local office of an international accounting firm—to undertake assessments of the financial management and administrative capacity of principal recipients, and to provide independent verification of substantive progress and financial accountability during the grant period. The board normally approves initial funding for two years (phase I), with approval for 3 additional years (phase II) being dependent on programme performance as evaluated by the secretariat near the end of phase I. This evaluation is the focus of our analysis.

There are three main components to the evaluation: (1) an appraisal of results, comparing services delivered and progress achieved with targets specified in the grant agreement; (2) a compilation of in-country LFA assessments completed during phase I on procurement, monitoring and evaluation, and progress toward stated goals; (3) a review of country contextual information such as key political cycles, conflict, natural disasters, etc.

On the basis of this information, the secretariat assigns one of four evaluation scores, which it makes publicly available on its website in a grant scorecard (table 1).^{3,4} "A" is awarded to those programmes meeting or exceeding performance expectations; "B1" to those with adequate performance; "B2" for inadequate performance but with potential demonstrated; and "C" for unacceptably poor performance. The board then decides whether to continue the funding with few or no programme changes; continue funding if the principal recipient meets certain

Lancet 2007; 369: 1807–13

See [Comment](#) page 1768

Center for Global Development,
Washington DC, USA
(S Radelet PhD, B Siddiqi MPhil)

Correspondence to:
Steven Radelet,
Center for Global Development,
1776 Massachusetts Avenue,
NW, Washington DC 20036, USA
sradelet@cgdev.org

	A: Meeting or exceeding expectations	B1: Adequate	B2: Inadequate but potential shown	C: Unacceptable
Number of people reached with services	Targets met or exceeding 80%	Significant improvements made (50–80%)	Some improvements made (30–50%)	Marginal or no improvements made (<30%)
Number of service centres established or strengthened	If the programme has achieved significant improvements in terms of numbers of persons reached, the Global Fund does not need to consider lower-level indicators for the Phase 2 decision		Significant improvements made (>30%)	Marginal or no improvements made (<30%)
Number of people trained to deliver services			Significant improvements made (>30%)	Marginal or no improvements made (<30%)

Source: Reference 3. Results are actual results as compared to targets for key coverage indicators.

Table 1: Global Fund programme evaluation scores

conditions; continue funding only if targets and budgets are substantially modified; or discontinue the grant.

Our aim in this analysis was to identify characteristics of the programme (eg, target disease, type of principal recipient), the recipient country's health sector (eg, number of doctors, donor crowding), and the country itself (eg, income, whether a government is socialist or not) that correlate with evaluation scores. We also aimed to investigate the association between changes in those characteristics and the probability of receiving particular evaluation scores. Our research builds on the secretariat's initial assessments of evaluation scores by examining a much wider set of characteristics and by analysing them in a more rigorous multivariate framework.^{1,3} It also builds on other research that has focused on Global Fund disbursements rather than evaluation scores.^{5–8}

Methods

We obtained data from the Global Fund on evaluation scores for the first 140 programmes evaluated in 2006 for phase II renewal. We excluded five grants with multi-country recipients (which made it difficult to assess health sector and country characteristics), and a grant to Somalia because of insufficient supplementary data. The remaining 134 grants included 28 (21%) with an A grade, 72 (54%) with a B1, 28 (21%) with a B2, and six (4%) with a C. The C category was too small to analyse independently, so we combined B2 and C grants into a single group of 34 grants.

We explored the association between evaluation scores and characteristics of the programme, the health sector, and the country. Programme characteristics include the target disease; the grant size per head; the type of recipient (government, multilateral, or NGO/private sector agency); the review panel's assessment of the original proposal; the round in which the Global Fund approved the grant; the organisation that acts as the LFA; and programme complexity, among others. Health sector characteristics included the number of doctors per head; the immunisation rate (either measles or DTP3 [diphtheria, tetanus, and pertussis]); the target disease prevalence rate; public health spending; and "donor crowding", measured for instance by the ratio of the size of the Global Fund grant to related donor funding. Country characteristics included income level; economic stability

(eg, the rate of inflation and the budget deficit); quality of governance; control of corruption; whether a country was categorised as a fragile state; the extent of civil liberties; political instability; and whether a country has or has had a socialist government.

Grant characteristics

For each grant, we collated the two-year phase I grant commitment, the amount disbursed so far, the target disease (HIV/AIDS, tuberculosis, malaria, HIV/tuberculosis, or integrated); the type of principal recipient (civil society or private sector; government; or multilateral), the disbursement round (1, 2, or 3), and the LFA (Price Waterhouse Coopers [PWC], KPMG, etc). In addition, the number of "service delivery areas" specified in each grant scorecard or grant performance report was used as a proxy for grant complexity.

Four measures of grant size were calculated for each grant: the dollar amount, the amount per person in the recipient country, the amount per person infected with the target disease in the recipient country, and the amount as a share of the recipient country's income. We recorded qualitatively similar results with each measure. The number of people infected in each country was calculated using WHO disease prevalence rates in each country. Country income and population data were derived from the World Bank's World Development Indicators.⁹

The technical quality of the grant was measured by previous grant proposal evaluations by the technical review panel, which categorises each successful proposal as either category 1 ("recommended with no or minor clarifications") or category 2 ("recommended contingent on clarifications met within a limited timeframe").

To assess the quality of the recipient country's CCM, we used several measures based on key responses to a 2005 study by the Technical Evaluation Reference Group (TERG) that evaluated 83 CCMs on their procedure and participation.¹⁰

Health sector characteristics

Data for the number of doctors per head, life expectancy, infant mortality and under-5 mortality were derived from the World Bank's World Development Indicators.⁹ Public health expenditure (% GDP) was compiled from World

Bank and national sources. We used WHO data for measles and DTP3 immunisation rates and disease prevalence rates—all numbers were for 2003 (the initial year for most grants under review) or the nearest year available. Disease prevalence was tallied using WHO data for the specific disease the grant was targeting. HIV/AIDS prevalence data were available only in broad ranges, so we used the midpoint of the range. For HIV/tuberculosis or integrated grants we used the larger number.

We measured donor crowding using aid and donor data from the Organization for Economic Cooperation and Development (OECD). Our main indicator was the ratio of the Global Fund grant size to total donor funding categorised by the OECD as used for “STD control including HIV/AIDS” and “infectious disease control” (the OECD has no data for aid used specifically for tuberculosis or malaria). As alternative measures, we assessed the ratio of the Global Fund grant to all health sector and population aid, and a simple count of the number of donors active in the health sector. For all these measures, we used data on commitments rather than disbursements. While disbursements might be preferable as a truer measure of “donor crowding,” they are, to some extent, a performance indicator. Grant disbursements are likely to be slower for poorly performing grants, so would be implicit in grant performance. Thus, to be consistent we used OECD commitments data. As a check, we measured the variable using actual disbursements and found similar results

Country characteristics

We compiled data for two definitions of weak states: the UK Department for International Development (DfID)’s definition of “fragile states” and the World Bank’s definition of Low Income Countries Under Stress (LICUS). Additional measures of institutional quality were also compiled, including: CPIA (Country Policy and Institutional Assessment) ratings, extracted from publicly available data for the 2003 CPIA quintiles released by the World Bank; six governance indicators from the World Bank Institute; and Political Rights and Civil Liberties indicators compiled by Freedom House.

Business cost numbers come from the World Bank’s Doing Business database. The IMF’s World Economic Outlook database was used for budget deficit numbers. Inflation data come from the IMF’s International Financial Statistics, data from recipient country central banks, and IMF Article IV reports. Data on adult literacy and girls’ primary school completion come from the World Bank. Finally, data on internal conflict come from the International Country Risk Guide; we averaged the conflict score over 2002–03.

Statistical analysis

We estimated the indices of an equation in which the evaluation score (converted from A, B1, and B2/C to 3, 2,

and 1, respectively) was the dependent variable; and programme, health sector, and country characteristics were independent variables. We estimated the relation between characteristics and score using an ordered multinomial probit methodology (with the STATA *oprobit* command), which estimates the probability of a programme receiving a particular rating given an array of characteristics. This methodology allowed us to estimate the relation between a marginal change in the independent variables and the probability of a programme receiving a particular evaluation score.

We note that this methodology allows us to show associations between grant characteristics and evaluation scores but not necessarily causality. We used robust cluster analysis to calculate the standard errors of the estimated coefficient, since the evaluation scores of different grants in the same country could be correlated, and we could not assume that the residuals of our estimation were independent (this method also corrects for the effect of heteroscedasticity and influential observations).

McFadden’s R^2 or Likelihood Ratio index compares the likelihood of an intercept-only model to the likelihood of the model with predictors; McKelvey and Zavoina’s R^2 measures the proportion of variance of the latent variable accounted for by the model; and the Count R^2 , which is simply the proportion of ratings correctly classified by the model.

In the webappendix we discuss these techniques in depth. Estimation results using different techniques and controlling for different combinations of independent variables are available from the authors.

See Online for webappendix

Role of the funding source

The funder of the study had no role in study design; data collection, analysis, or interpretation; or writing of the report. The corresponding author had full access to all the data and had final responsibility for the decision to submit for publication.

Results

Table 2 shows our main results. Programmes in which a government was the principal recipient received significantly lower scores than those with civil society, private sector, or multilateral recipients. After controlling for other characteristics, a programme with a government recipient was 16·7% less likely to receive an A grade, as likely as programmes with other recipients to receive a B1, and 16·8% more likely to receive a B2/C.

Evaluation scores tended to fall as grant size (measured on a per head basis) increased, suggesting the possibility of slight capacity constraints with larger grants. Earlier work identified a similar association between grant size and disbursement speed.⁵ The relation was non-linear (as captured by the log of the grant per head): each additional dollar was associated with a lower evaluation score, but the marginal effect diminished as the grant size

For more on governance indicators from the World Bank Institute see www.worldbank.org/wbi/governance

For more on political rights and civil liberties indicators see www.freedomhouse.org

For the Doing Business database see www.doingbusiness.org

For more on the International Country Risk Guide see www.icrgonline.com

increased. An increase of one SD in the log of the grant size per head was associated with an 8·8% decrease in the probability of receiving an A score, and a 10·9% increase in the probability of receiving a B2/C score. We also noted that evaluation scores for malaria programmes were significantly lower than for HIV/AIDS or TB programmes. For example, malaria programmes were 12·9% less likely than HIV/AIDS or TB programmes to receive an A.

Programmes approved by the board in the first round tended to receive lower scores than those approved in the second or third rounds. With respect to the quality of an original proposal, the technical review panel makes one of four recommendations: accept, accept with revisions, revise and resubmit, and reject. All approved Global Fund programmes receive one of the first two ratings. Programmes that received an “accept” recommendation were 17·6% more likely to receive an A grade and 16·8% less likely to receive a B2/C grade than those that received an “accept with revision” recommendation.

The accounting firm KPMG was the LFA in 43 sample programmes, PWC in 71, and other firms in 20. Programmes in which KPMG is the LFA tended to have

lower evaluation scores: even after controlling for other variables, only 12% of these grants received an A rating, whereas 25% of PWC and other grants received an A. There were no significant associations between other grant characteristics, such as programme complexity (measured by the number of “service delivery areas”) and assessment scores of the quality of CCM operations with evaluation scores.¹⁰

Programmes in countries with more doctors per head had significantly higher evaluation scores. A one standard-deviation increase in the number of physicians per head (equivalent to just over one more physician per 1000 people) is associated with a 10·2% higher probability of receiving an A and a 12·6% lower probability of receiving a B2/C (p=0·002). Countries with higher measles immunisation rates tended to receive higher evaluation scores. We identified qualitatively similar results for the DTP3 immunisation rate, but with a larger SD for the estimated coefficient.

Evaluation scores tended to be lower in countries with more donors, and higher where there were few. Specifically, in countries where the Global Fund grant forms a large portion of total donor funding for sexually transmitted diseases (STDs) and infectious diseases, evaluation scores tended to be higher. We explored several variations of this measure, including a denominator with just STD donor funding, a denominator with all donor health funding, and a simple count of donors active in STDs and infectious diseases, and found qualitatively similar results for each measure.

Evaluation scores were slightly higher in countries with higher prevalence rates for the target disease of the programme. The result was not strongly robust, and sometimes lost statistical significance in alternative specifications. None of a wide range of other health characteristics, including life expectancy, infant mortality, government public health spending as a share of GDP, and births attended by skilled health staff, were significantly associated with evaluation scores after controlling for other variables.

Income levels were negatively related to evaluation scores, after controlling for other variables. Other analyses have found similar results, which are the opposite of what might be expected.¹⁵ Specifically, a one standard-deviation increase in the log of per head income was associated with a 7·6% drop in the probability of receiving an A grade, and a 9·4% increase in the probability of receiving a B2/C.

We also examined grant scores in politically fragile states. In terms of simple averages, programmes in countries categorised by the UK’s DfID as fragile states tended to receive lower scores. However, after controlling for other variables these differences disappeared, in particular after controlling for the number of physicians per head and immunisation rates, both of which tend to be lower in fragile states (see the webappendix). Programmes in countries that were or had been (ie,

	Regression coefficients (SE)	Association between the marginal change in variable and the change in probability of receiving an evaluation score			
		Magnitude of change	Change in probability (SE)		
			A	B1	B2/C
Government principal recipient*	-0.745 (0.275)†	From 0 to 1	-16.7% (6.7%)‡	0.0% (4.1%)	16.8% (5.7%)†
Ln Grant per capita	-0.293 (0.098)†	1 SD	-8.8% (2.7%)†	-2.1% (2.8%)	10.9% (4.2%)†
Malaria grant*	-0.796 (0.219)†	From 0 to 1	-12.6% (3.7%)†	-10.8% (5.8%)§	23.3% (7.4%)†
High technical review panel rating*	0.764 (0.306)‡	From 0 to 1	17.6% (8.1%)‡	-0.8% (4.1%)	-16.8% (7.0%)‡
Disbursement round	0.518 (0.279)§	1 SD	6.5% (3.5%)§	1.6% (2.2%)	-8.0% (4.8%)§
LFA was KPMG*	-0.627 (0.265)‡	From 0 to 1	-11.0% (4.7%)‡	-6.2% (4.5%)	17.2% (7.6%)‡
Measles immunisation rate	0.015 (0.008)§	1 SD	5.0% (3.3%)§	1.7% (1.7%)	-6.7% (3.3%)§
Physicians per 1000 people	0.458 (0.148)†	1 SD	10.2% (3.9%)†	2.4% (2.8%)	-12.6% (4.2%)†
Ln Global Fund share of infectious diseases and STD aid	0.515 (0.191)†	1 SD	7.9% (3.0%)†	1.9% (2.4%)	-9.9% (3.9%)†
Ln Disease prevalence rate	0.150 (0.084)§	1 SD	5.8% (3.3%)§	1.4% (1.9%)	-7.2% (4.3%)§
Ln GDP per capita	-0.401 (0.170)‡	1 SD	-7.6% (3.6%)‡	-1.9% (2.2%)	9.4% (4.1%)‡
3-year budget deficit	-0.104 (0.039)†	1 SD	-6.4% (2.5%)†	-1.5% (1.8%)	8.0% (3.1%)‡
Socialist in 1990*	0.606 (0.332)§	From 0 to 1	14.4% (8.5%)§	-1.6% (4.1%)	-12.8% (6.1%)‡

Evaluation scores: A=3, B1=2, B2/C=1. *yes=1, †p<0.01, ‡p=0.05, §p=0.10. Measures of fit: McKelvey and Zavoina’s R²=0.537; McFadden’s R² (Likelihood Ratio index)=0.282; Count R² (proportion of accurate predictions)=0.657. The interpretation of the coefficient in an ordered probit model is not straightforward (see webappendix), so in columns 2–5 we report the marginal effects of changes in the predictors on the probability of receiving a particular rating. So, for a given magnitude of change (column 2) in an explanatory variable (specifically, 0 to 1 for dummy variables, and one standard deviation for continuous variables), columns 3, 4 and 5 report the associated change in probability of receiving an A, B1, or B2/C score, respectively. Ln=natural log. Data from alternative estimation models are available from the authors.

Table 2: Evaluation scores and programme, health sector, and country characteristics for 134 programmes

around 1990) socialist tended to have higher evaluation scores. This result was not significant in some specifications, but was always positive. Programmes in countries with larger government budget deficits tended to receive lower scores.

With respect to other characteristics, evaluation scores were not linked to any of the World Bank Institute's six measures of governance (including corruption and political stability);¹¹ Freedom House's measures of civil liberties and political rights; whether a country suffered from internal conflict; the World Bank's CPIA score; or measures of "red tape", such as the number of days required to start a business. We also found no relation with adult literacy rates or girls' primary school completion rates. Finally, country geographical characteristics (landlocked or located in the tropics) were not associated with differences in evaluation scores.

Discussion

Global Fund programmes that scored lowest had government agencies as principal recipients, received a large amount of funding, were focused on malaria, had weak initial proposals, or had KPMG as the LFA. Countries with a high number of doctors per head, high measles immunisation rates, few health sector donors, and high disease prevalence rates had higher evaluation scores. Poor countries, those with small government budget deficits, and those that have or have had socialist governments also received higher scores.

Our finding that programmes with government principal recipients receive weaker evaluation scores is consistent with earlier analyses on disbursement speed.^{1,5-7} This result could partly be due to differences in programmatic focus or procurement systems across recipients, or because government recipients are working with a larger number of donors and hence have larger administrative burdens. But bureaucratic and capacity problems in government programmes are a more likely answer, suggesting that NGO and civil society groups have a better capacity to implement programmes effectively. This result does not suggest that the Global Fund should have a bias against programmes with government recipients, but rather that it should encourage countries to facilitate programmes with non-government actors alongside government programmes.

The weaker malaria programme evaluation scores, which have been noted by the Global Fund,³ might be due partly to the switch to artemisinin combination therapy for drug-resistant malaria in 2004, which slowed implementation and added to costs. If so, malaria programme performance could improve once the new protocol is fully introduced.

Our finding that evaluation scores differ depending on the LFA even after controlling for other characteristics suggests that the Global Fund can strengthen its evaluation system by achieving more consistency across LFAs. We cannot say from this analysis whether KPMG

or PWC, or both, should adjust their approach, but it seems apparent that the two are using different standards in evaluating programmes. As part of its upcoming 5-year review, the Global Fund is analysing and evaluating the LFA system, which provides a good opportunity to strengthen this component of the evaluation system.

More programmes in which the original proposal received an "accept" recommendation received an A grade than those with an "accept with revision" recommendation, suggesting that the technical review panel, to some extent, identifies stronger programmes. But it is also possible that programmes with a strong evaluation could be perceived as "better" programmes from the outset, which could bias evaluators when they assign their subsequent evaluation scores.

Some observers have speculated that programmes approved in the Global Fund's first round were weaker than subsequent ones, since the Fund's systems were not initially fully in place. We identified a slight tendency for programmes approved in later rounds to receive higher evaluation scores. This could indicate an improvement in the systems but could also be because of timing—it is likely that the best programmes from any round are evaluated first. Our sample contains only 15% of the round three grants, and, as more are evaluated, the average round three score is likely to diminish, and with it the difference in average evaluation scores across rounds.

Evaluation scores were slightly higher in countries with higher prevalence rates for the target disease of the programme. This result could indicate greater public awareness and stronger commitment to fight the diseases. Alternatively, it might simply be easier to achieve measurable progress in countries with higher prevalence, or there could be systematic differences in how targets are set in high prevalence countries. The strong association between the number of physicians per head and evaluation scores, coupled with the positive association with immunisation rates, suggests that countries with stronger health systems and larger numbers of trained health workers are more likely to have successful programmes. (Earlier studies have identified a strong link between health sector institutional characteristics and measles and DTP3 immunisation rates.¹²) These findings suggest the need for greater Global Fund and CCM oversight in countries with weaker systems and capacity. But they also underscore the importance of building strong health systems to combat the three diseases more broadly rather than focusing narrowly on short-term targets.

There are several possible reasons why evaluation scores were weaker in programmes where there were many other donors. The administrative and management demands on recipients are much greater when there are multiple donors which could weaken performance. It may be that the incentives for strong performance are weaker when recipients have many funding alternatives.

Alternatively, the presence of many other donors could elicit more public criticism of Global Fund programmes (justified or not, some of which might come from competing donors), which could influence the perceptions of evaluators. We cannot at this stage disentangle which of these might be the case in different countries. It does suggest the Global Fund and other donors should consider streamlining their administrative burdens in these countries and harmonising systems where possible. For example, the Global Fund provides financing to Mozambique through a basket funding approach in conjunction with other donors. The Fund could also focus more of its efforts in countries where there are fewer donors, consistent with its role of filling funding gaps.

Evaluation scores tend to be inversely related to income levels, even after controlling for other characteristics. It is difficult to interpret this result, since it could be due to any number of factors correlated with incomes that we have inadvertently excluded. However, it does show there is no reason to expect that low-income countries will necessarily perform poorly.⁶ Being resource-poor might weaken implementation capacity through the impact of fewer physicians and a lower immunisation rate but, after accounting for these factors, evaluation scores do not suffer further.

The reason that programmes in countries with larger government budget deficits received lower evaluation scores is not obvious, but one possibility is that a larger deficit could indicate generally weaker government economic and financial management. Weaker oversight systems could both weaken the overall financial stance and the ability to implement programmes. But this interpretation is conjecture, as more direct measures of financial management are not available.

We noted that programmes in fragile states received lower evaluation scores, but that the relationship disappeared once we controlled for the number of doctors and immunisation rates. Thus, it is correct to conclude that programmes in fragile states on average tend to receive weaker scores, but this difference can be accounted for by the number of physicians, immunisation rates, and other factors controlled for in the analysis. Similarly, scores were higher in socialist countries, suggesting that countries formerly in the Soviet bloc have stronger health systems and implementation capacity, even after accounting for the number of physicians and immunisation rates.

The Global Fund's evaluation system has strengths and weaknesses, and we recognise that the evaluation score may not fully reflect true programme performance. The evaluation system includes certain subjective elements, which, while appropriately allowing for flexibility when relevant contextual information is not captured by the numbers, can also introduce inconsistencies across programmes and over time. Different Global Fund portfolio managers or LFAs might weigh contextual

information differently, and assessments can be influenced by personal biases, the competence of the assessors, or political or bureaucratic pressure to give higher scores to certain programmes. Moreover, accurately judging performance is difficult because programmes have different degrees of difficulty, and it will be easier to receive an A in programmes with easily met targets than in those with more ambitious ones. Thus, it is possible that a score of A on one grant might not actually imply stronger "performance" in a real sense than a score of B1 on another, any more than a school grade of A from one teacher might not signify more learning than a B grade from another teacher. Of course, nearly every evaluation system faces these difficulties to some extent, and the Global Fund continually tries to strengthen its system. Thus, our work should be seen as an analysis of evaluation scores rather than of the actual performance of the programmes.

It is important to note that our methodology shows associations between characteristics and evaluation scores but not necessarily causality. Reverse causality is not a major concern, as it is hard to see how evaluation scores could change the characteristics we examine. However the associations we detect could be due to an unmeasured third factor that is related to both the characteristics and the grant score. In some cases, we strongly suspect causality, but we cannot at this stage prove it. In addition, care must be taken in extrapolating our results outside of the sample, since there might be some selection bias in the programmes that the Global Fund approves or among those that were evaluated first.

Nevertheless, these results provide insights that could help the Global Fund manage its risks by better predicting which programmes are likely to receive higher evaluation scores. We do not suggest that the Fund uses this analysis to decide which proposals to finance or how to distribute its funding, but rather as information to assess the risks on approved programmes and to guide allocation of its management resources for oversight and risk management. The results also suggest indicators the Secretariat can use for its early warning systems. We recognise that these results are not definitive, but can offer initial guidance until stronger data and analytic tools are developed.

This initial analysis of evaluation scores is far from perfect, but provides a good starting point that can be improved as the Global Fund strengthens the measurement of both performance and the key factors that are likely to influence performance over time.

Contributors

SR conceived the idea, led the analysis, and wrote most of the text. BS contributed to the model development, collected most of the data, and conducted the primary data analysis. Both authors read and approved the final version.

Conflict of interest statement

The Global Fund paid for travel expenses only for SR to attend one conference in 2005. We declare that we have no conflict of interest.

Acknowledgments

We thank Celina Schocken for her valuable input to this analysis, four anonymous referees for comments on an earlier draft, and Bernhard Schwartlander and Daniel Low-Beer from the Global Fund for help with the data (though they bear no responsibility for this analysis). We also thank the William and Flora Hewlett Foundation for their generous financial support of this work. This study was carried out independently of the Global Fund, and all opinions and errors are those of the authors.

References

- 1 The Global Fund to Fight AIDS, Tuberculosis and Malaria. Investing in impact: mid-year results report 2006. Geneva: GFATM; 2006.
- 2 The Global Fund to Fight AIDS, Tuberculosis and Malaria. Current grant commitments and disbursements. http://www.theglobalfund.org/en/funds_raised/commitments/ (accessed December, 2006).
- 3 The Global Fund to Fight AIDS, Tuberculosis and Malaria. Sustaining performance, building up results: third progress report 2005. Geneva: GFATM; 2005.
- 4 The Global Fund to Fight AIDS, Tuberculosis and Malaria. Phase II: grant scorecards. http://www.theglobalfund.org/en/funds_raised/gsc/ (accessed December, 2006).
- 5 Lu C, Michaud CM, Khan K, Murray CJL. Absorptive capacity and disbursements by the Global Fund to Fight AIDS, Tuberculosis and Malaria: analysis of grant implementation. *Lancet* 2006; **368**: 483–88.
- 6 AIDSPAN. Analysis: Global Fund Grants are three months behind schedule, on average. Global Fund Observer Newsletter (Geneva), May 5, 2005. <http://www.aidspace.org/documents/gfo//GFO-Issue-45.htm> (accessed December, 2006).
- 7 Radelet S. The Global Fund to Fight AIDS, Tuberculosis, and Malaria: progress, potential, and challenges for the future. Washington: Center for Global Development; 2004.
- 8 US Government Accountability Office. The Global Fund to Fight AIDS, TB, and Malaria is responding to challenges but needs better information and documentation for performance-based funding. Washington: US Government Accountability Office; 2005. <http://www.gao.gov/new.items/d05639.pdf> (accessed December, 2006).
- 9 World Bank. World Development Indicators 2006. Washington: World Bank; 2006.
- 10 Technical Evaluation Reference Group. Report on the assessment of country coordinating mechanisms: performance baseline. Geneva: GFATM; 2005. http://www.theglobalfund.org/en/about/terg/announcements/assessment_161205/default.asp (accessed December, 2006).
- 11 World Bank. Worldwide Governance Indicators 2006. <http://web.worldbank.org/WBSITE/EXTERNAL/WBI/EXTWBI/GOVANTCOR/0,,contentMDK:20771165~menuPK:1866365~pagePK:64168445~piPK:64168309~theSitePK:1740530,00.html> (accessed December, 2006).
- 12 Khaleghian P, Gauri V. Immunisation in developing countries: its political and organizational determinants. *World Development* 2002; **30**: 2109–32.