**METHODS BRIEF**

PCE Impact Evaluation Methodological Summary

David E Phillips PhD

This is a draft and not intended for distribution. March 28, 2018

# Purpose of this Document

The purpose of this document is twofold:

* 1. To explain, in advance, the analytical approach which we have selected for impact evaluation, with the goals of methods improvement through feed- back, transparency, and formal demonstration of the robustness (both strengths and limitations) of the methods. There are many aspects of this ap- proach which I believe are as robust as possible, although the overall study design is weak.
  2. To lay out a “road map” of how we can accomplish a task as daunting as an impact evaluation of the Global Fund in 2 years. To detail (more) precisely how this can be done.

# Overview

The basic approach to this impact evaluation is simply to measure many separate indicators, then measure their correlation. Through impact evaluation frame- works (i.e. “results chains”) and thematic frameworks, those indicators (and more) are outlined in specific detail. The remainder of the impact evaluation is to identify the appropriate indicators measure correlation between, and do so while controlling for confounding (see *Worked Example* section below).

The majority of the complexity and effort required to accomplish this task is to ensure that the data we use actually produce a valid measure of the quantity of interest (e.g. that case notifications reflect popula- tion incidence). Techniques include data triangulation (see Section 9), model-based geostatistics (see Section 10), outlier detection, and multiple imputation (see Section 8), will be applied selectively to enhance the

Correspondence: [davidp6@uw.edu](mailto:davidp6@uw.edu)

Institute for Health Metrics and Evaluation, University of Washington, Seattle, USA

measurement validity. Special attention is given to re- source tracking to accurately measure inputs (see *Fur- ther Details on Resource Tracking* ).

The controls are especially critical for the impact analysis (see Section 3). The basic approach of “mea- suring many indicators” also includes measuring co- variates and controls that are not necessarily the indi- cators of interest, but are essential the measuring the correlation between expenditure and outputs.

Finally, it is essential that the indicators are mea- sured independently. That is to say that the measure- ment approach for a specific indicator does not use any of its preceding indicators (in the results chain) in the process. This will help ensure that the correla- tions measured between indicators are reflective of the theorized causal pathways, not endogeneity in mea- surement.

# Hypothesis Being Tested

The core hypothesis of this impact evaluation is that changes in Global Fund investments result in observ- able changes in health systems outputs. Additionally, the hypothesis is that those changes in outputs result in observable changes in intervention coverage, which subsequently results in improvements in burden of dis- ease.

We assume that, conditional on need, Global Fund investments are exogenous to health systems out- puts. In other words, we assume that investments are targeted towards interventions and areas with un- met need, and that apart from decision-making about where there is need for investment, health systems out- puts are only affected by Global Fund investment and do not affect it. Need, in this context is defined as burden of disease without proportionate spending by other financing sources.

# Basic Outputs Model

The central model for measuring the contribution of Global Fund inputs to health systems outputs is a lin- ear model. The precise functional form of the model will be selected to be fit-for-purpose given the distri- bution of the output data. In theory, the character- istics of typical output data are that they are inte- ger (whole numbers), over-disbursed (requiring both

a mean and variance parameter to describe their dis- tribution), heteroskedastic (different variance at dif- ferent levels of a given explanatory variable), and temporally-autocorrelated (similar to itself in previ- ous time points). For these reasons, the model is (ar- guably) most appropriately fit as a generalized linear mixed model (GLMM) in the negative binomial fam- ily. For simplicity, I will avoid model notation that explicitly represents a negative binomial GLMM, in order to focus attention on the variables and the level at which they’re measured (i.e. indexed).

4.1 Extension to RSSH

In evaluating the contribution of catalytic and system- wide Global Fund investments (such as investments to strengthen resilient and sustainable systems for health (RSSH)), an additional layer of controls may be necessary. This is because such investments are intended to operate in addition to, or synergistically with, other program areas, not to result in outputs on their own.[CITE] Essentially, this amounts to control- ling for Global Fund spending from programs other than just RSSH as well:

Measurement of the Global Fund contribution to outputs will rely on the co-variance of outputs and ex-

*Ojti* = *β*1*E*

*RSSH*

*jti*

+ *β*1*E*

*HIV/T B/Malaria*

*jti*

(2)

penditure along three dimensions: space, time and ac-

+*β*2*E¬gf*

3 *jt*

tivity. In other words, investments by the Global Fund

*jti* + *β I*

(3)

focus on different interventions in different places, and change from year to year, and so do outputs. If the hypothesis being tested is correct, we would expect to observe changes in health systems outputs that co- incide with changes in investments along those three dimensions (space, time and activity). As noted al- ready, the coincidence between changes in outputs and changes in investments alone is not expected to reflect the Global Fund’s contribution, but rather the coinci- dence that happened apart from the correlation that other resources and burden of disease have with out- puts. That implies the following regression:

*Ojti* = *β*1*Egf* + *β*2*E¬gf* + *β*3*Ijt* (1)

*Technical Note: Although expenditure from the Global*

*Fund is the intervention of interest, the above model will also be explored using the approved budget as a proxy for expenditure. The rationale for doing so is to avoid potential bias from differential absorption. Similar to the “intent to treat” principle from a ran- domized control trial, the drivers of actual execution of funds are more diverse and poorly-understood than the drivers of planned budgeting, and thus more chal- lenging to control. To measure the correlation between actual expenditure and outputs would also be to mea- sure the contribution of those absorption-driving fac- tors, and would weaken the interpretability of the coef- ficient of interest. While both budget and expenditure will be explored, budget will be considered the primary explanatory variable.*

*jti*

*jti*

In this formula, *j* indexes subnational areas (provinces/ **5 Relationship with Value for Money**

departments, districts/municipalities or otherwise), *t* indexes time (years, quarters, or months) and *i* in- dexes intervention categories, as defined by the Global Fund’s Modular Framework. *O* represents counts of health systems outputs, *E* represents expenditure ei- ther by the Global Fund (*gf* ) or other sources (*¬gf* ), and *I* represents incidence of the disease for which interventions were designed. The *β* terms are coeffi- cients (correlations) to be estimated from the data us- ing maximum likelihood estimation (or similar). This

The basic outputs model has a natural relationship

with value for money (VfM) assessment. As defined by the Global Fund Monitoring and Evaluation team, the key metric of VfM is cost per output, or cost per case averted. [CITE]

The definition of the coefficient *β*1 in formula 1 is the average observed increase in outputs per unit increase in expenditure. *β*1 is also simply stated as the inverse of cost per output (i.e. *cost per output* = 1 ). In this

*β*

1

way, the core model for impact evaluation is the same

model will primarily be fit separately by disease.

The coefficient *β*1 therefore measures the correlation between Global Fund investment and outputs, control- ling for other investments and disease burden. Under the assumptions of the PCE Theory of Change and Impact Frameworks (detailed elsewhere), this correla- tion represents the contribution of the Global Fund to those outputs.

as the core model for VfM assessment.

The key to using the impact evaluation model for VfM assessment is additional interpretation of it. As allocation efficiency is among the critical topics for the PCE, estimates from the core model can be produced among counterfactual scenarios, comparing alternate mixes of interventions and their expected relationship with output.

Table 1: Simulated Example Data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | **Exp.** | **Exp.** |  |
| **Intervention** | **Area** | **Year** | **Indicator** | **Inc.** | **(Non-GF)** | **(GF)** | **Output** |
| Differentiated HIV testing services | 1 | 2012 | HIV tests conducted | 384 | 95 | 33 | 1143 |
| Differentiated HIV testing services | 1 | 2013 | HIV tests conducted | 355 | 131 | 27 | 1111 |
| Differentiated HIV testing services | 1 | 2014 | HIV tests conducted | 279 | 107 | 20 | 1027 |
| Differentiated HIV testing services | 1 | 2015 | HIV tests conducted | 290 | 99 | 52 | 1051 |
| Differentiated HIV testing services | 2 | 2012 | HIV tests conducted | 508 | 168 | 46 | 1267 |
| Differentiated HIV testing services | 2 | 2013 | HIV tests conducted | 429 | 142 | 79 | 1197 |
| Differentiated HIV testing services | 2 | 2014 | HIV tests conducted | 412 | 106 | 59 | 1171 |
| Differentiated HIV testing services | 2 | 2015 | HIV tests conducted | 368 | 2 | 77 | 1100 |
| Differentiated HIV testing services | 3 | 2012 | HIV tests conducted | 539 | 163 | 66 | 1297 |
| Differentiated HIV testing services | 3 | 2013 | HIV tests conducted | 467 | 114 | 41 | 1225 |
| Differentiated HIV testing services | 3 | 2014 | HIV tests conducted | 463 | 109 | 36 | 1207 |
| Differentiated HIV testing services | 3 | 2015 | HIV tests conducted | 385 | 30 | 87 | 1126 |
| Differentiated HIV testing services | 4 | 2012 | HIV tests conducted | 564 | 119 | 72 | 1311 |
| Differentiated HIV testing services | 4 | 2013 | HIV tests conducted | 475 | 138 | 50 | 1237 |
| Differentiated HIV testing services | 4 | 2014 | HIV tests conducted | 508 | 188 | 34 | 1272 |
| Differentiated HIV testing services | 4 | 2015 | HIV tests conducted | 429 | 103 | 44 | 1178 |
| Differentiated ART service delivery | 1 | 2012 | ARV doses administered | 384 | 113 | 28 | 1131 |
| Differentiated ART service delivery | 1 | 2013 | ARV doses administered | 355 | 86 | 43 | 1104 |
| Differentiated ART service delivery | 1 | 2014 | ARV doses administered | 279 | 8 | 89 | 1026 |
| Differentiated ART service delivery | 1 | 2015 | ARV doses administered | 290 | 54 | 15 | 1027 |
| Differentiated ART service delivery | 2 | 2012 | ARV doses administered | 508 | 109 | 84 | 1265 |
| Differentiated ART service delivery | 2 | 2013 | ARV doses administered | 429 | 105 | 62 | 1171 |
| Differentiated ART service delivery | 2 | 2014 | ARV doses administered | 412 | 153 | 48 | 1181 |
| Differentiated ART service delivery | 2 | 2015 | ARV doses administered | 368 | 126 | 23 | 1113 |
| Differentiated ART service delivery | 3 | 2012 | ARV doses administered | 539 | 127 | 78 | 1291 |
| Differentiated ART service delivery | 3 | 2013 | ARV doses administered | 467 | 105 | 69 | 1215 |
| Differentiated ART service delivery | 3 | 2014 | ARV doses administered | 463 | 147 | 31 | 1214 |
| Differentiated ART service delivery | 3 | 2015 | ARV doses administered | 385 | 121 | 60 | 1147 |
| Differentiated ART service delivery | 4 | 2012 | ARV doses administered | 564 | 110 | 109 | 1304 |
| Differentiated ART service delivery | 4 | 2013 | ARV doses administered | 475 | 87 | 89 | 1235 |
| Differentiated ART service delivery | 4 | 2014 | ARV doses administered | 508 | 143 | 87 | 1259 |
| Differentiated ART service delivery | 4 | 2015 | ARV doses administered | 429 | 142 | 48 | 1187 |
| Male circumcision | 1 | 2012 | Circumcisions administered | 384 | 91 | 25 | 1121 |
| Male circumcision | 1 | 2013 | Circumcisions administered | 355 | 128 | 13 | 1106 |
| Male circumcision | 1 | 2014 | Circumcisions administered | 279 | 88 | 1 | 1043 |
| Male circumcision | 1 | 2015 | Circumcisions administered | 290 | 45 | 41 | 1035 |
| Male circumcision | 2 | 2012 | Circumcisions administered | 508 | 142 | 43 | 1251 |
| Male circumcision | 2 | 2013 | Circumcisions administered | 429 | 56 | 80 | 1161 |
| Male circumcision | 2 | 2014 | Circumcisions administered | 412 | 167 | 1 | 1175 |
| Male circumcision | 2 | 2015 | Circumcisions administered | 368 | 181 | 5 | 1137 |
| Male circumcision | 3 | 2012 | Circumcisions administered | 539 | 118 | 63 | 1284 |
| Male circumcision | 3 | 2013 | Circumcisions administered | 467 | 70 | 117 | 1214 |
| Male circumcision | 3 | 2014 | Circumcisions administered | 463 | 141 | 59 | 1213 |
| Male circumcision | 3 | 2015 | Circumcisions administered | 385 | 90 | 69 | 1133 |
| Male circumcision | 4 | 2012 | Circumcisions administered | 564 | 249 | 24 | 1346 |
| Male circumcision | 4 | 2013 | Circumcisions administered | 475 | 117 | 94 | 1231 |
| Male circumcision | 4 | 2014 | Circumcisions administered | 508 | 158 | 35 | 1253 |
| Male circumcision | 4 | 2015 | Circumcisions administered | 429 | 109 | 44 | 1194 |

1

# 6 Worked Example

This section provides an example of the model above, using simulated (i.e. completely fabricated) data. It is intended to provide a tangible demonstration of the analysis, not (necessarily) to reflect real numbers per- fectly.

Table 1 displays some example data. This simu- lated dataset contains information on three interven- tions: Differentiated HIV testing services, Differen- tiated ART service delivery and Male circumcision. These are HIV interventions selected from the Modu- lar Framework. The table depicts these interventions

exist between the two sources of expenditure, and be- tween expenditure and incidence, since Global Fund investments are intended to be coordinated with other spending, and both are intended to target areas of greatest need. The consequence of this is that, con- trolling for incidence (or expenditure) yields a less- confounded correlation between expenditure and out- put, as shown in Figure 2.

**Figure 2** Correlation Stratified by Incidence Quantile

Quantile 1 Quantile 2 Quantile 3 Quantile 4

●

●

●

●

●

●

●

implemented in four hypothetical subnational areas (numbered 1-4), which can be thought of as districts, municipalities, etc. Each of these intervention-areas has annual measurements from 2012 to 2015. Finally, each intervention is labeled with an associated output indicator: HIV tests conducted, ARV doses adminis- tered and Circumcisions administered. An “estimate” of HIV incidence is displayed in the fifth column. Note that incidence does not vary between interventions,

2500

2000

Output (Units)

● ●

1500

●

●

●

●

Area

● 1

● 2

● 3

● 4

only area-years. The final three columns in Table 1 display simulated expenditure from non-Global Fund sources, expenditure from the Global Fund, and Out- put (units of the corresponding indicator consumed).

●

●

●

These data have been constructed in such a way that Global Fund expenditure is positively correlated with outputs from each intervention, according to the the- orized data generating process. Figure 1 demonstates the correlation.

**Figure 1** Example Correlation between Output and Expenditure

0 25 50 75 0 25 50 75 50 75 100 20 40 60 80 100

Global Fund Expenditure (Million USD)

Here, the correlation between expenditure and out- put has been binned by quantiles of incidence. From this figure, the strong correlation between incidence and output is clear, as the higher quantiles have gen- erally higher output. The other result of displaying the correlation in this way however is a tighter rela- tionship between expenditure and output within each quantile. While it’s entirely possible that the opposite effect could be true with real data (i.e. a weaker corre- lation after controlling for incidence), the relationship

2500

2000

Output (Units)

Area

● 1

● 2

● 3

● 4

Intervention ●

Differentiated ART service delivery

* Differentiated HIV testing services ●

Male circumcision ●

●

●

●

●

● ●

●

●

between incidence and expenditure is likely to be im- portant for the primary correlation of interest.

As a result of these characteristics of the data and theory, a regression that controls for each of the afore- mentioned correlates of output is necessary, as dis- played in equation 1 above. Fitting an ordinary nega- tive binomial regression (in this case not GLMM, al-

●

●

●

1500

●

though it was mentioned above) yeilds the estimates of correlations shown in Table 1.

●

**Table 1 Regression Results**

Estimate Std. Error z value Pr(*>|*z*|*)

0 30 60 90 120

Global Fund Expenditure (Million USD)

(Intercept) 6.5838 0.0186 353.19 0.0000

Non-Global Fund expenditure, and even more so in- cidence also have positive correlations with outputs. Importantly, strong correlations are also expected to

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Expenditure gf | 0.0015 | 0.0002 | 7.82 | 0.0000 |
| Expenditure nongf | 0.0005 | 0.0001 | 3.74 | 0.0002 |
| Incidence | 0.0021 | 0.0001 | 28.80 | 0.0000 |

In other words, per dollar increase in Global Fund expenditure, outputs increase on average *e*0*.*0015, or

1.002 units as an average across areas, years and in- terventions (once other expenditure and incidence are taken into account). In this hypothetical data, this is a stronger correlation than other expenditure, but weaker than incidence.

1. **Further Details on Resource Tracking** The objective of resource tracking is to produce the most valid possible estimate of expenditure (both Global Fund and otherwise) by area, time and inter- vention. The basic problem is that while there are data sources that track resources, they typically only offer high detail on one or two of those three dimensions, re- quiring other data for further detail. Two approaches are being considered to combine these data sources. Each has benefits and drawbacks.
   1. Purely Systematic Model

This approach deterministically multiplies the inter- vention fractions from one data source by the area- module counts from another:

*Eti*

well, defined by the geographic regions where inter- ventions are operated by sub-recipients. In this way, the fraction in equation 4 can be computed specific to broad geographic areas at the least, allowing a more geographically-precise apportioning of interven- tions within modules. Another limitation of this ap- proach is data availability. It is unlikely that national- level data will cover the entire time period spanned by global-level data, leaving gaps in the ultimate esti- mates of *Ejti*.

* 1. Stochastic, Cost-Based Model

This approach uses a statistical model to predict ex- penditure using supply-chain data and costing esti- mates, treating resource tracking data as predictor variables:

*Djti* ∗ *C*--*i* = *β*1*Ejtm* + *β*2*Eti* + *E* (5)

As in the previous section, *Ejtm* is a quantity that can be measured from country-level data sources, and *Eti* can be measured from global-level data sources. The two of these can be thought of as strong predic-

*E*---*jti* = *Ejtm* ∗

*E*

*tm*

(4)

tors of the distribution of commodities *Djti*, which is measurable through other data sources and can also

Where *E* represents investment from any source (*E*ˆ being an estimate of *E*) and *m* indexes modules (from the Modular Framework), within which interventions are nested (in a collectively-exhaustive way). The frac-

be indexed by intervention *i*. By multiplying the es- timated unit cost of a commodity (*C*--*i*) by the dis- tribution of that commodity, the fitted values from regression formula are an estimator of the quantity of

tion *Eti*

*E*

*tm*

is computed *by module*, such that the sum

interest, *E*---*jti*.

of this fraction across the interventions of a single

module is %100 (), *Etim*∈*M*

*E*

*tm*=*M*

= 1). The purpose of

This approach has limitations as well. For one, we

only have *D* for a subset of all interventions (and prob-

this approach is to utilize data sources with different

strengths in combination. The PCE Inception Phase uncovered that certain country-level data sources of- ten contain a high level of detail about subnational ar- eas (*j*), but little detail about interventions, typically only the corresponding modules (*m*). Global-level data sources however often offer a high level of detail about interventions (*i*), but little, if any, detail about sub- national areas. Hence, the fraction in equation 4 can be accurately measured using global-level data, while *Ejtm* can be accurately measured using country-level data. By combining the two, we estimate the details of interventions within subnational areas.

This approach relies on the assumption that invest- ment is allocated with the same proportions within each module at every subnational area. This is likely untrue. One approach toward relaxing this assump- tion is to attempt to break down the fraction *Eti*

*E*

*tm*

(i.e. the global-level data) into subnational areas as

ably always will). To work around this, we would fit the model and then use out-of-sample prediction to esti- mate *E*ˆ for all intervention categories. Doing so would rely on the assumption that the relationship between expenditure and distribution for observable interven- tions is reflective of the relationship for unobserved interventions, which may not be true. However, the addition of more covariates in equation 5 could relax this assumption.

# Further Details on Data Processing

[Outliers] [Imputation]

1. **Further Details on Data Triangulation** [basic explanation of how to use two or more data sources in combination to produce a stronger estimate of a quantity that they each measure]

# Further Details on Model-Based Geostatistics

[basic explanation of spatial autocorrelation]

# References

[to fill in]