

**DORIS DUKE CHARITABLE FOUNDATION
PHIT PARTNERSHIP IMPLEMENTATION RESEARCH FRAMEWORK**

**Nirmala Ravishankar
Emmanuela Gakidou
Stephen Lim
Ziad Obermeyer
Christopher JL Murray**

**The Institute for Health Metrics and Evaluation
February 17, 2008**

This document was developed by the Institute for Health Metrics and Evaluation at the University of Washington to provide a common framework for assessing the activities of the Population Health Implementation and Training (PHIT) Partnerships funded through the Doris Duke Charitable Foundation's African Health Initiative (AHI).

AHI was launched in 2007 to help catalyze a shift from disease-specific programs to programs that provide integrated primary healthcare on a large scale and strengthen health systems. Each PHIT Partnership will focus on the following goals:

- Achieve measurable, significant health improvements by providing sustainable integrated primary healthcare to a substantial underserved region (a minimum of 250,000 people);
- Strengthen health systems and the health workforce in the region of interest; and
- Increase the knowledge for evidence-based health delivery and health systems planning by supporting implementation research.

AHI reflects DDCF's dual commitment to improving health outcomes and strengthening health systems in Africa, on the one hand, and contributing to the global evidence-base on successful healthcare delivery through implementation research on the other. In the context of AHI, the term "implementation research" refers to the generation of knowledge about the delivery of healthcare to target populations under real-world conditions.

The emphasis on implementation research addresses what has come to be recognized in the public health literature as a gap between growing knowledge about the prevention and cure of diseases and the small evidence-base for effective ways of delivering medical treatments to those who need them in developing countries. Even while we continue to develop and test innovations such as new vaccines, drugs, and medical procedures through rigorous clinical research, we know relatively little about how best to deliver them to poor people in countries with weak health systems. In the absence of scientific research on healthcare delivery and program implementation, existing treatments and innovations known to have positive effects on health outcomes in clinical settings often fail to achieve high coverage in real-world settings in developing countries.¹ Even for public health interventions that leading experts believe to be successes, the evidence-base when it comes to documented impact on health outcomes remains weak at best.² This knowledge gap persists because, as a recent report by a group of experts reviewing evaluation practices found, governments and implementing agencies "do well in their normal data collection and evaluation tasks related to monitoring inputs, improving operations, and assessing performance, but largely fail in building knowledge. . ."³

The PHIT Partnerships, each of which will be designing and implementing a model for delivering integrated primary healthcare, offer an invaluable opportunity for undertaking

¹ For a recent discussion about the science of implementation research, see Madon et al. (2007).

² Savedoff & Levine (2004).

³ Evaluation Gap Working Group (2006), p. 2.

scientific implementation research. Having in place a system for undertaking such research will serve multiple goals. First, it will help the PHIT Partnerships identify and address gaps in their programs, facilitating greater learning and innovation. Second, it offers a way to benchmark improvements in health outcomes and health system strengthening achieved by the PHIT Partnerships. Finally, assessing the results of different integrated primary healthcare models is essential if the PHIT Partnerships are to generate generalizable findings that contribute to the global evidence base about how best to deliver integrated primary health care.

This document develops a common conceptual framework for implementation research for the purposes of AHI. It is meant to serve as a starting point for designing specific research strategies tailored to the needs of each PHIT Partnership. The PHIT Partnerships will develop these strategies over the course of the planning grant period, in collaboration with DDCF and an external evaluation group. Having such a team approach to implementation research ensures that the research strategy maximizes opportunities for learning and continued improvement over the course of the program, while meeting the standards of scientific research.

Implementation research in this framework has the following three components:

- Program evaluation, which measures the impact of the PHIT Partnership on health outcomes and generates generalizable findings through counterfactual analysis, i.e. by comparing changes in program areas with those in control areas;
- Economic evaluation, which gauges the economic efficiency or the cost-effectiveness of the program; and
- Monitoring and operations research, which focuses on assessing program inputs and processes, tracking health system outputs, and undertaking research on specific elements of the program through targeted analysis.

The first section of the document presents a typology for understanding the different dimensions of the PHIT Partnership programs. The second section describes in more detail the three components of implementation research listed above. The third section discusses study design, while the fourth section focuses on health and health system indicators commonly used in implementation research on health service delivery.

A. The Object of Study: The PHIT Partnership

The first step in designing an implementation research strategy entails identifying and articulating the particular intervention or program that is being assessed. In each of the PHIT Partnerships, the intervention is a complex and dynamic primary health care delivery program that brings together a package of health services and a system for delivering it. In other words, the object of study is not a single medical treatment or even a package of treatments, but instead a model for delivering a complex package of services to an underserved community. The goal of implementation research is to assess the strengths and weaknesses of that model.

At a conceptual level, there are four functional dimensions to any health system: provision of health services, resource generation, financing, and stewardship.⁴ While the exact content of the integrated primary healthcare model will vary across the PHIT Partnerships, each program will likely touch on some or all of these dimensions. The four dimensions, as they apply to the PHIT Partnerships, are described below:

1. **Provision of Health Services:** refers to the package of health services or treatments that a health system delivers. In the context of AHI, the emphasis on integrated primary health care implies that the package will likely be comprehensive while the exact composition of the package will vary based on local needs and priorities. For example, the relative emphasis on communicable diseases versus non-communicable diseases will differ depending on the needs of the target population. The set of potential health services can be categorized into the following broad categories:
 - i. **Personal health services:** refers to services that are delivered to patients during a direct interaction with a health provider
 - ii. **Non-personal health services:** refers to actions that are either applied at the community level, like health education and healthy living promotion campaigns, or relate to non-human components of the environment, like sanitation

Annex B provides an example of a package of health services including both personal and non-personal interventions. Tables in Annex C present background information on the burden of disease and risk factors and a ranking of interventions by effectiveness for sub-Saharan Africa.

2. **Resource generation:** refers to the host of inputs that go into a health system. PHIT Partnerships may allocate resources to build and strengthen capacity within the existing health system. These could include investments of physical resources like health facilities, mobile units, diagnostic tools, and equipment; human resources, both in terms of the number of health workers and their level of training; drugs; and information systems like telecommunication and computers.
3. **Financing:** refers to the way funds are managed within a health system. PHIT Partnership may explore one or more health financing mechanisms affecting both health care providers and users. For example, the way the provider payment system is structured creates different incentives for health care providers that in turn have important implications for the quantity and quality of health services. On the demand side, conditional cash transfers programs create financial incentives for users to seek care.
4. **Stewardship:** refers to the way a health system is managed and regulated. One of the central goals of AHI is to foster innovation in this area, which includes new strategies for increasing community participation, community-directed treatments,

⁴ This framework for conceptualizing the different functions of health systems and assessing their performance was developed by Murray & Frenk (2000).

new norms and standards, referral systems, and regulation and oversight systems for health providers. PHIT Partnerships may explore different stewardship models.

B. Components of Implementation Research

Implementation research has come to mean different things in different fields. For the purposes of AHI, implementation research is conceptualized as having the following three components:

1. **Program Evaluation:** The main goal of program evaluation is to determine what impact, if any, the program had on the target population.⁵ To do so, it relies on a comparison between what happened to the target population receiving the program and a suitable comparison group that is similar to the group receiving the program in every way except for access to the program. Studies that compare outcomes in program areas with non-program areas without taking into account systematic differences between the two groups will lead to biased estimates of the program effect. Hence, baseline measurement of performance indicators is needed to check that the comparison group is in fact similar to those receiving the program and to take any observable differences between the two groups into account while estimating the effect of the program. Additionally, isolating the effect of the program of interest is necessary for arriving at an accurate measure of the effect size. Evaluation studies that are based on a comparison of outcomes before and after a program was introduced for the same group fall short of this goal since there is no way to rule out the possibility that some other policy or event that coincided with the program caused the observed change in outcomes. Therefore, the comparison group should be distinct from the treatment group. These two principles—a) pre- and post-intervention measurement of outcomes and b) non-overlapping program and comparison groups—are widely accepted as best practices in program evaluation.⁶
2. **Economic Assessment:** While the impact of an integrated primary health program on health outcomes is of interest by itself, the efficiency with which a program is improving health or the cost-effectiveness⁷ of the program is an important performance indicator, especially in Africa where the financial resources available for health are limited. Hence, equal attention must be paid to measuring health care costs as is paid to assessing health outcomes. In the

⁵ The Evaluation Working Group and the Global Health Policy Network, both of which are hosted by the Center for Global Development (www.cgdev.org), offer a wealth of information on program evaluation in global health, including an overview of best practices and common challenges, and in-depth case studies of recent health programs and their evaluation strategies. The World Bank's online resource on impact evaluation available at www.worldbank.org/impactevaluation provides links to documentation on evaluation methods and a searchable database of World Bank-funded evaluations of anti-poverty programs.

⁶ The Evaluation Working Group (2006).

⁷ For more information on cost effectiveness analysis, see Drummond et al. (1997) & Adam et al. (2005).

context of integrated primary healthcare programs implemented by the PHIT Partnerships, the focus of the economic evaluation should be on the net difference in costs and benefits between the program and control groups, e.g. the incremental cost per Disability Adjusted Life Year (DALY) averted. As the program may incur both additional costs related to innovations of the program as well as reduced health care costs downstream, measurement of health care costs should be made for both program and control groups and should include all costs that can be expected to vary between the groups. Additionally, measurement should be based on the concept of economic or opportunity cost rather than financial cost. For example, the cost of health care volunteer time should be included and valued appropriately. The perspective used for measuring cost data (e.g. health system or societal) should be specified upfront and should at the very least aim to collect both public and private expenditure directly related to health services.

3. **Monitoring and Operations Research:** The goal of monitoring is to ascertain whether the program is being implemented according to plan and identify problem areas. Operations research, on the other hand, focuses on specific elements of the program and provides a more targeted assessment. While similar to program evaluation in its use of comparison groups, operations research provides insights into particular components of a program, like the use of provider payments or a particular workforce training module. Together, program monitoring and operations research facilitate feedback and learning during the course of the program. Monitoring systems that track program inputs, processes, and outputs highlight areas where the program could be strengthened. Similarly, operations research can provide timely assessments of specific program elements, which then could be used to initiate change. They both typically rely on a combination of quantitative and qualitative methods to maximize the breadth and depth of learning. Quantitative methods are suitable for tracking systemic changes and population metrics. Qualitative methods, such as focus groups and open-ended interviews, allow for more nuanced research on smaller sub-groups, like health providers or health professionals, and on sensitive research topics.

C. Study Design

Counterfactual analysis, defined by a comparison of what happened to a population receiving a program with what would have happened to the same population had they not received the program, is a feature of most implementation research and lies at the heart of program evaluation in particular. Since the counterfactual pertains to a hypothetical situation, i.e. we do not observe what the outcomes would have been in the absence of the program once the program has been introduced, this comparison cannot be made in practice. Consequently, implementation science relies on comparisons between the program group and a suitable comparison group that is similar to the program group in every way except for access to the program.

Ensuring that the program and comparison groups are truly comparable at baseline such that any change in the post-intervention phase can be attributed to the program can pose significant challenges. They are discussed below.

Confounding

When receiving the program is associated with other factors that also have an independent effect on program outcomes, then the measured effect of the program is distorted by these confounding factors. In other words, confounding factors result in the program and the comparison group not being similar in every way except for program status. For example, suppose the program evaluation of a PHIT Partnership was done by comparing health outcomes for those who voluntarily avail of care offered by the program with those who do not. Those who seek care are likely to be composed of people who are less healthy than the group that does not participate. Upon receiving the care, their health may improve, but they may still be less healthy than those who did not avail of the new program. A simple comparison of the two groups—those who received care under the new program and those who did not – would lead us to conclude that the program had a negative effect on health. Health prior to the introduction of the program, which is associated with health-seeking behavior and post-intervention health, is the confounding factor in this example.

Any number of individual and household characteristics that make certain people more likely to participate in the program than others and are also linked to their health (e.g. age, sex, education, household income) can be potential confounding factors in a study. Features of particular communities, villages, administrative districts, and health facilities can also be potential confounders. In the context of AHI, the introduction of other programs in either the program or control area during the course of the evaluation could lead to confounding.

Broadly speaking, there are two ways to address the problem of confounding: randomization-based methods and non-experimental statistical methods.

- i. Randomization-based experimental methods⁸: Randomization is the best way to address the problem of confounding. The ideal situation would be if we could randomly select two large samples of individuals in a country and assign one to receive the program, while withholding the program from the other group. As long as the samples are large and selected randomly, members of the two groups will on average be similar in every way except for receiving the program. Consequently, any difference in outcomes between the two groups can be directly attributed to the program.

Individual randomization, which is frequently used in clinical research, is not always feasible for evaluating community-level programs. In the context of AHI, it would be difficult to offer an integrated health care program to some members

⁸ Gertler (2000), King, et al. (2007), and Miguel & Kremer (2004) are examples of how experimental methods based on randomized assignment have been used to evaluate health interventions and public health programs.

of a household or a few households in a village and not others for practical, political, and ethical reasons. Cluster randomization, which refers to random assignment of the program at a higher level of aggregation like a village, administrative district, or health facility, is often easier to achieve with health programs. Randomly assigning some clusters to receive the program takes care of the problem of confounding at that level. However, if the sample of clusters is small, which is often the case in cluster randomized studies of health programs, randomization may not solve the problem of confounding. Hence, the applicability of cluster randomization to evaluating the PHIT Partnership activities depends on the level of the health system that the program targets. If the program is at the level of local health centers or dispensaries and there are numerous potential program units within the target area, then randomization would be an appropriate way for reducing confounding. However, if the program involves investing in the district hospital and there are only a handful of hospitals in the target region, then randomization will not address the problem of confounding. Finally, cluster randomization may not always control for individual confounders. Hence, statistical methods discussed below are often used in addition to randomization to arrive at the best estimate of the program effect.

- ii. Non-randomized statistical methods: When randomization is not an option or does not fully address the problem of potential confounders at different levels, then quasi-experimental statistical methods are used to correct for measurable confounders. These statistical methods, which can often be used in conjunction with each other and in addition to randomization, are:
 - a. Difference in difference design⁹: this method involves comparing changes before and after the program for individuals in the program and control groups. The regression-based model attempts to address the problem of confounding by controlling for difference at baseline.
 - b. Matching¹⁰: this method matches program users from the program group with one or more non users from the control group based on their background characteristics, thereby controlling for any confounding due to those factors. A variant of this method uses an individual's propensity or probability of participating in the program to do the matching, where the propensity scores are estimated as a function of individual characteristics.
 - c. Instrumental Variables¹¹: an instrumental variable is a variable that is correlated with program participation but is unrelated to other background characteristics and health outcomes of interest. Such a variable can be used to independently estimate program participation. Next, this estimate of program participation that is independent of potential confounders is used to estimate the effect of the program on health outcomes of interest. For example, suppose a random sub-sample of the families in the program area is selected to receive additional information about the program in the form of a door-to-door outreach campaign. Receiving additional information is likely to lead to

⁹ For application of this method, see Galiani S et al. (2005) & Newman et al. (2002).

¹⁰ For application of this method, see Galido & Briceno (2004) and Wagstaff & Yu (2005).

¹¹ For application of this method, see Ruiz-Arranz et al. (2002) and Attanasio & Vera-Hernandez (2004).

participation in the program but is unrelated to other family characteristics and health since the selection for the outreach program was done randomly.

- d. Regression discontinuity¹²: this method is applicable if a cut-off point, like a poverty threshold, is used to select program participants. Families that straddle the cut-off point are similar except some receive the program, while others do not. The regression discontinuity method derives its estimate of the program effect by comparing outcomes for families just below and above the arbitrary threshold for eligibility.

Effect Size Modification

Even if an evaluation study design takes potential confounding into account using the methods described above, there are additional factors that can modify the size of effect that is measured. The main reasons for why this might occur in the current context are discussed below:

- i. Low adherence: Not everyone in the program regions will avail of health care even if they need it. On the flip side, some in the comparison group may receive care through the primary health care program. The estimated effect of a program derived from comparing outcomes between a control group and program group will vary based on the extent of adherence to assignment in the two groups. For instance, some families in the program areas may not find out about the new primary health care program and consequently not seek care. Similar problems of non-adherence to the assignment can arise if the community receiving the new integrated primary health care program is adjacent to the community that has been selected as the control area. It is likely that some from the control area will cross over to the program group, in turn changing the size of the effect measured. It is typical in the fields of medicine and public health to estimate the effect of the program by measuring the intention to treat (ITT) effect, which should be the norm here.
- ii. Variable Program Effect: Implementing the program in the real world means that the actual program may differ from the program as it was intended. Moreover, the program may vary from one health unit to another. In the case of the PHIT Partnerships, the primary health care program offered in different health facilities may vary in quality due to both systematic factors and random differences. This internal variation in the program is typically unavoidable and will cause the size of the effect to vary. One primary goal of monitoring is to document these differences.
- iii. Externalities: The PHIT Partnership could have unanticipated positive or negative consequences for those outside the program area. An example of a positive externality is when a PHIT Partnership generates demand for better health systems in non-program areas, which in turn leads to the adoption of similar models elsewhere. If, instead, the program draws away trained medical

¹² For application of this method, see Levy & Ohls (2007) and Canton & Blom (2004).

professionals from other areas, then it would lead to a negative externality. Factoring such unintended consequences into implementation research is essential for determining the true effects of the program.

D. Indicators for Implementation Research

DDCF will be funding up to 6 PHIT Partnerships to design and implement integrated primary health care delivery models in sub-Saharan Africa. The Partnerships will be undertaking implementation research to measure the extent to which each of the programs improve health outcomes in the target community and strengthen the health system. Changes in health outcomes, however, typically take several years to manifest themselves. Consequently, interim measures of health system performance will also be needed to benchmark the achievements of the PHIT Partnerships.

It is conventional to classify performance indicators into the following three broad categories: (1) health outcomes, (2) program outputs, and (3) program inputs and processes. A typology of indicators that are relevant to AHI is presented below along with a description of the indicators that fall within each category:

1. Health outcome measures: refers to the three main goals of a health system, namely health, fairness of contribution, and responsiveness.
 - i. Population health outcomes¹³: Improving the health of the target population is the defining goal of a health care program. Metrics for measuring population health include:
 - a. Child mortality: Under-1 and under-5 mortality
 - b. Adult mortality: Age- and sex-specific mortality rates, as well as a summary measure of adult mortality such as 45q15, i.e. the probability of dying between the ages of 15 and 59.
 - c. Causes of death: Numbers of deaths attributable to the major causes. The list of major causes might vary slightly across countries, but will likely have significant overlap. The composition of the leading 10 causes of death for children and adults should be monitored as useful input into the epidemiologic profile of the population.
 - d. Disease-specific health outcomes and risk factors. These should be decided separately for each program, depending on the composition of the package of services being delivered.
 - ii. Health expenditure¹⁴: This is measured in terms of catastrophic health spending and out-of-pocket expenditure. Indicators include:
 - a. Total amount of health expenditure from all sources
 - b. Amount of out-of-pocket health expenditure
 - c. The proportion of households that spend more than 30% of their disposable income on health

¹³ Refer to chapter 26 in Murray & Evans (2003).

¹⁴ For more information, refer to chapters 38-42 in Murray & Evans (2003).

- d. The proportion of health expenditure that results from out-of-pocket payments (as opposed to being financed through pre-payment mechanisms)
- iii. Responsiveness¹⁵: Responsiveness captures the non-medical aspects of the interaction between a patient and the health system. Indicators of the responsiveness of health systems are critical to measure during the implementation of a new system of delivering health care. For the purposes of AHI, the relevant indicators for measuring the responsiveness of outpatient and inpatient services include:
- a. Quality of care, including the cleanliness of the facilities, the quality and cleanliness of the patient beds, the availability of food during inpatient stay, patient satisfaction etc.
 - b. Promptness of care/Waiting time, such as average waiting times in facilities and average waiting times to get specialized care, when needed.
 - c. Access to social networks (mostly for inpatient care), such as whether patients are able to have their family members and other members of their social network visit during their hospital stay.
 - d. Communication between providers and patients, such as whether diagnoses are effectively communicated to the patient, and whether the patient understands what they are supposed to do upon leaving the facility in terms of taking medication, follow-up visits, etc.
2. Program output measures¹⁶: These are measures of the direct output of the health system; they can change in a very short period of time and any change in them can be directly attributed to the health system. Therefore, they can be used for monitoring progress throughout the implementation of the program and identifying areas of weakness in the program, and for evaluating the impact of the program.
- i. Coverage: for the set of interventions that are being delivered through a program, coverage is defined as the proportion of the population receiving an intervention out of all those in need of the intervention. In other words, it measures the number of people who received an intervention (the numerator) out of the universe in need of the intervention (denominator). Coverage is measured separately for each intervention and then aggregated into a composite measure of health system coverage.
 - ii. Effective coverage: Effective coverage takes into consideration the quality of the intervention being delivered. Quality ranges from 0 to 1; if the individual receiving the intervention gets the maximum health gain from it then quality equals 1. If an intervention is being delivered but it results in no health gain to an individual, then quality equals 0. Measures of effective coverage are important to monitor as they track both the

¹⁵ For more information, refer to chapters 43-47 in Murray & Evans (2003).

¹⁶ For more information, refer to Shengelia et al. (2005), Lozano et al. (2006), and chapter 20 in Murray & Evans (2003).

population receiving interventions but also the quality of the interventions being delivered.

3. Health system inputs and process measures¹⁷: these refer to resources invested in the health system and activities introduced to achieve program goals. Indicators in this category track the following:
 - i. Human resources, such as measures of health personnel per 1,000 people, number of personnel completing training per year, new recruits, attrition rates etc.
 - ii. Infrastructure and equipment, such as complete inventories of buildings, technological and laboratory equipment available
 - iii. Drug supply, including the types and quantities of drugs available in the area of intervention and broken down by districts/sub-areas (where relevant)
 - iv. Operational measures, including how many hours per day and how many days per week the facilities are providing services, measures of the management of the referral system, etc.
 - v. Program activities, such as number and types of community outreach programs, educational materials and workshops for the population, etc.

E. Measurement and Data Collection

While the study design for undertaking implementation research is likely to vary between PHIT Partnerships, the requirements for data collection, data quality and data management are similar. These common themes are discussed below.

Sources of Data

During the planning phase, it is in the interest of the PHIT Partnerships to collect data from all existing sources in order to design integrated primary healthcare programs that respond to local needs, on the one hand, and implementation research strategies that have the power to detect the effect of their program on the other.

Once the program commences, the PHIT Partnerships will need a multi-prong approach to data collection given both the complexity of the programs they will be implementing and the rarity of some health conditions. The most relevant sources of data for the purposes of AHI are:

- i. Population Health Surveys: A population-based health survey that is administered in program and control areas should serve as the primary source for information on individual-level health system output and outcomes. In designing the survey the following factors should be considered:
 - a. Survey content: the population survey should measure household characteristics (including socio-economic indicators, size of the household, and proximity to a health facility), individual health

¹⁷ For more information, refer to chapter 24 in Murray & Evans (2003).

characteristics (including gender-specific risk factors) and health outcomes using biomarkers.

- b. Sample size: each of the PHIT Partnership will be serving a target population of at least 250,000 people. The random sample drawn from this population should be large enough to detect the effect of the program on relatively rare conditions. Hence, the Partnerships should use power calculations to determine the appropriate sample size.
 - c. Sample design: the need for clustered, multi-stage, or stratified sampling will depend on the target population. An overlapping panel/cross-section sample design would be desirable for the purposes of implementation research. The panel design will allow for inferences about changes over time while avoiding survivor bias. The repeated cross-sections will address the problem of program attrition and preempt any incentives that a panel design might create for program administrators to solely target households in the sample.
- ii. Health Facility Survey: refers to information from health units on health operations, consultations, emergencies, personnel, infrastructure, equipment, and drug inventories.
 - iii. Administrative Data: this includes data on program costs as well as data from provider registries, hospital records, hospital and health facility financial records, and disease surveillance sites.

Timing of Data Collection

Specific decisions regarding the frequency of data collection and the types of data collected in each cycle will depend on several factors, including the needs of the monitoring system and ongoing operations research, the cost of collecting data from different sources, and expectations of when changes in the outcomes of interest are likely to occur. The following template for data collection in the implementation phase takes these competing considerations into account:

- i. Annual collection of administrative data in both program and control areas
- ii. Periodic collection of data from household surveys and facility surveys, including:
 - a. Baseline measurement of all indicators in program and control areas before the program is introduced
 - b. Interim measurement in treated areas after 2 years of operation, focusing on process measures, output measures and outcome measures that are likely to have changed during this time (for example: child mortality, certain disease-specific health outcomes, catastrophic health expenditure and out-of-pocket spending)
 - c. Final measurement in both program and control areas in the final year of the project

Rolling surveys of households, if feasible, could replace the interim and final population surveys.

Sources of Errors

Ensuring good quality data poses several challenges. The main sources of bias during the survey design and data collection phase of the study are:

- i. Selection bias: refers to errors in the measurement of the program effect caused by the selection of an unrepresentative sample from the target population.
- ii. Measurement bias: refers to errors caused by inaccurate measurement of variables.

Anticipating these sources of error and taking measures to address them can minimize these types of biases. These include data validation, field testing survey instruments, adequate supervisor and investigator training, and additional protocols for data management.

Confidentiality of Information

Much of the data that the PHIT partnerships will collect in order to undertake implementation research is likely to pose concerns regarding confidentiality. The Partnerships are expected to follow standards set by institutional review boards overseeing research at the US-based institution as well as standards set by the program country. This process should be initiated as early in the program planning process as possible.

Conclusion

Funding implementation research on the delivery of integrated primary healthcare is one of the key goals of DDCF's African Health Initiative. Each PHIT Partnership will be implementing a model for delivering integrated primary healthcare and designing an implementation research strategy for their program. Research of this kind is needed to both benchmark what the PHIT Partnerships have achieved in terms of the stated goals of improving health outcomes in the target population and strengthening the health system, and for generating generalizable findings that can contribute to the global evidence-base on the delivery of integrated primary healthcare.

Works Cited

- Adam T, Manzi F, Schellenberg JA, Mgalula L, de Savigny D, Evans DB. Does the integrated management of childhood illness cost more than routine care? Results from the United Republic of Tanzania. *Bull World Health Organ* 2005;83:369-77.
- Attanasio O, Vera-Hernandez AM. Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia. *Working Paper EWP04/06*, London: Centre for the Evaluation of Development Policies, Institute of Fiscal Studies, 2004.
- Canton E, Blom A. Can student loans improve accessibility to higher education and student performance? An impact study of the case of SOFES, Mexico. *Policy Research Working Paper No. 3425*, Washington, DC: World Bank, 2004.
- Evaluation Gap Working Group. When Will We Ever Learn? Improving Lives Through Impact Evaluation. Washington DC: *Center for Global Development*, 2006.
- Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes, 2nd Edition*. Oxford: Oxford University Press, 1997.
- Galiani S, Gertler PJ, Schargrodsky E. Water for life: the impact of the privatization of water services on child mortality. *J Polit Econ* 2005;113: 83-120.
- Galido V, Briceno, B. Evaluating the impact on child mortality of a water supply and sewerage expansion in Quito: is water enough? *OVE Working Paper 0105*, Washington, DC: Inter-American Development Bank, 2004.
- Gertler PJ. Final report: the impact of PROGRESA on health. Washington, DC: International Food Policy Research Institute, 2000.
- King G, Gakidou E, Ravishankar N, Moore RT, Lakin J, Vargas M, Téllez-Rojo MM, Hernández Ávila JE, Hernández Ávila M, Hernández Llamas, H. A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program. *J Policy Anal Manage* 2007;26:479-506.
- Levy D, Ohls J. Jamaica: Evaluation of Jamaica's PATH Program: Final Report. Washington, DC: *Mathematica Policy Research, Inc.*, 2007.
- Lozano R, Soliz P, Gakidou E, Abbott-Klafter J, Feehan DM, Vidal C, Ortiz JP, Murray CJL. Benchmarking of performance of Mexican states with effective coverage. *Lancet* 2006;368:1729-41.

- Madon T, Hofman K, Kupfer L and Glass R, "Implementation Science", *Science*, December 17, 2007.
- Miguel E, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 2004;72:159-217.
- Murray CJL, Frenk J. A framework for assessing the performance of health systems. *Bull World Health Organ* 2000;78:717-31.
- Murray CJL, Evans DB, eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization, 2003.
- Newman J, Pradhan M, Rawlings LB, Ridder G, Coa R, Evia JL. Impact evaluation of social funds: an impact evaluation of education, health, and water supply investments by the Bolivian Social Investment Fund. *World Bank Econ Rev* 2002;16:241-74.
- Ruiz-Arranz M, Davis B, Stampini M, Winters P, Handa, S. More Calories or More Diversity? An Econometric Evaluation of the Impact of the PROGRESA and PROCAMPO transfer programs on food security in rural Mexico. Boston: Harvard University, 2002.
- Savedoff WD, Levine R, Learning from Development: the Case for an International Council to Catalyze Independent Impact Evaluations of Social Sector Interventions. Washington, DC: *Center for Global Development*, 2006. www.cgdev.org.
- Shengelia B, Tandon A, Adams OB, Murray CJL. Access, utilization, quality, and effective coverage: An integrated conceptual framework and measurement strategy. *Soc Sci & Med* 2005;61:97-109.
- Wagstaff A, Yu S. Do health sector reforms have their intended impacts? The World Bank's Health VIII Project in Gansu Province, China. *World Bank Policy Research Working Paper 3743*, Washington, DC: *World Bank*, 2005.

Annex A: Sample of a Monitoring & Evaluation Timeline

	Planning grant proposal	Planning Phase	Year 1	Year 2	Year 3	Year 4	Year 5	End of project
Identification of program region								
Preliminary outline of program								
Preliminary outline of study design and performance measurement								
Identification of existing data sources								
Assessment of availability (and possibly quality) of existing data								
Identify potential partners in area of interest for collaboration in data collection efforts								
Design of the treatment								
Collation of data from existing sources and assessment of quality								
Selection of treatment and control groups								
Design survey instruments								
Apply for IRB approval								
Baseline data collection								
Administrative data								
Health facility survey								
Household survey								
Interim measurements								
Administrative data								
Health facility survey								
Household survey								
Final data collection								
Administrative data								
Health facility survey								
Household survey								
Analysis of baseline data								
Analysis of administrative data								
Analysis of interim survey data								
Assessment of the final impact of the project								
Assessment of final cost-effectiveness of the program								

Annex B: Sample package of services

1. Child health
 - a. Immunizations (DPT, Polio, MMR, BCG, Hib, HBV, MCV, Rotavirus)
 - b. Breast-feeding and low-birth weight baby care education
 - c. Nutritional monitoring and treatment, micronutrient supplementation (vitamin A, zinc, iron only if setting-appropriate)
 - d. Treatment of common childhood diseases
 - i. Diarrhea (education for families including oral rehydration and triage, management at health facilities)
 - ii. ARI (education, case management)
 - iii. Malaria (free ITN distribution, case-management)

2. Women's Health
 - a. Family planning services
 - b. Cervical cancer screening, HPV vaccination
 - c. Antenatal care
 - i. maternal immunizations and nutritional supplements
 - ii. screening for preeclampsia, syphilis, HIV
 - iii. Treatment of HIV (PMTCT), malaria, etc
 - d. Obstetric and neonatal care
 - i. screening and referral system for obstetric emergencies
 - ii. facility-based emergency obstetric and emergency neonatal care
 - iii. management of pre-term births

3. Prevention, Detection and Treatment of Communicable Diseases among Adults
 - a. HIV/AIDS (voluntary counseling and testing, ARVs, education)
 - b. Malaria (ITNs, spraying, case management)
 - c. Pulmonary infections (diagnosis and management) including TB (chemotherapy for smear-positive, -negative, and drug-resistant cases, active case finding)
 - d. Other communicable diseases (e.g. STIs, helminthic infections, leishmaniasis, trypanosomiasis, trachoma, influenza, unsafe injections)

4. Prevention, Detection and Treatment of Non-Communicable Diseases and risk-factors
 - a. Cardiovascular disease: risk factor reduction and treatment (hypertension, lipids, etc)
 - b. Diabetes diagnosis and treatment
 - c. Smoking reduction through education campaigns
 - d. Injury prevention (e.g. helmet use) and emergency treatment
 - e. Diagnosis and treatment of neuropsychiatric conditions
 - f. Other non-communicable diseases and risk factors (e.g. surgical emergencies, cancer, micronutrient deficiencies, indoor air pollution, cataracts)

Annex C: Statistical tables

Table 1: DALYs by Cause, Sex, and Age in Sub-saharan Africa, 2001 (thousands)

Cause	Total	Male	Female
<i>Population (millions)</i>	668	331	336
All causes	344,754	175,141	169,613
I. Communicable, maternal, perinatal, and nutritional conditions	242,837	118,494	124,343
A. Infectious and parasitic diseases	173,484	86,027	87,458
1. Tuberculosis	8,084	5,350	2,734
2. Sexually transmitted diseases excluding HIV/AIDS	3,842	1,679	2,163
a. Syphilis	2,347	1,223	1,124
b. Chlamydia	559	52	507
c. Gonorrhoea	894	379	515
d. Other sexually transmitted diseases	41	24	17
3. HIV/AIDS	56,820	27,005	29,815
4. Diarrheal diseases	22,046	11,506	10,540
5. Childhood-cluster diseases	23,198	11,568	11,630
a. Pertussis	6,116	3,047	3,069
b. Poliomyelitis	17	8	9
c. Diphtheria	45	26	19
d. Measles	13,539	6,752	6,787
e. Tetanus	3,481	1,735	1,746
6. Meningitis	941	469	473
7. Hepatitis B ^a	536	301	235
Hepatitis C ^a	217	124	93
8. Malaria	35,447	16,801	18,646
9. Tropical-cluster diseases	4,897	3,277	1,620
a. Trypanosomiasis	1,310	830	480
b. Chagas' disease	—	—	—
c. Schistosomiasis	1,184	701	483
d. Leishmaniasis	312	223	89
e. Lymphatic filariasis	1,656	1,269	387
f. Onchocerciasis	436	255	181
10. Leprosy	24	14	9
11. Dengue	4	2	3
12. Japanese encephalitis	—	—	—
13. Trachoma	1,455	357	1,097
14. Intestinal nematode infections	905	452	453
a. Ascariasis	476	235	241
b. Trichuriasis	119	58	61
c. Hookworm disease	309	158	151
Other intestinal infections	1	1	1
Other infectious diseases	15,068	7,123	7,945
B. Respiratory infections	31,107	16,846	14,261
1. Lower respiratory infections	30,455	16,517	13,938
2. Upper respiratory infections	371	173	198
3. Otitis media	281	156	125
C. Maternal conditions	9,743	—	9,743
1. Maternal hemorrhage	1,643	—	1,643
2. Maternal sepsis	1,843	—	1,843
3. Hypertensive disorders of	842	—	842

pregnancy			
4. Obstructed labor	919	—	919
5. Abortion	1,557	—	1,557
Other maternal conditions	2,940	—	2,940
D. Perinatal conditions^b	20,047	11,351	8,696
1. Low birthweight	7,891	4,501	3,391
2. Birth asphyxia and birth trauma	9,256	5,195	4,062
Other perinatal conditions	2,899	1,655	1,244
E. Nutritional deficiencies	8,455	4,271	4,184
1. Protein-energy malnutrition	5,220	2,722	2,498
2. Iodine deficiency	951	487	464
3. Vitamin A deficiency	548	249	299
4. Iron-deficiency anemia	1,688	789	899
Other nutritional disorders	49	24	25
II. Noncommunicable diseases	73,069	35,829	37,240
A. Malignant neoplasms	6,281	3,092	3,189
1. Mouth and oropharynx cancers	284	193	91
2. Esophageal cancer	343	212	131
3. Stomach cancer	487	258	229
4. Colon and rectal cancers	291	158	133
5. Liver cancer	762	512	250
6. Pancreas cancer	117	52	64
7. Trachea, bronchus, and lung			
cancers	225	162	63
8. Melanoma and other skin cancers	118	53	64
9. Breast cancer	574	1	574
10. Cervix uteri cancer	627	—	627
11. Corpus uteri cancer	41	—	41
12. Ovarian cancer	152	—	152
13. Prostate cancer	416	416	—
14. Bladder cancer	133	88	45
15. Lymphomas and multiple myeloma	622	386	236
16. Leukemia	245	128	118
Other malignant neoplasms	844	472	372
B. Other neoplasms	188	95	93
C. Diabetes mellitus	1,448	547	901
D. Endocrine disorders	2,706	1,425	1,281
E. Neuropsychiatric conditions	15,151	7,593	7,558
1. Unipolar depressive disorders	3,275	1,291	1,985
2. Bipolar affective disorder	1,204	615	590
3. Schizophrenia	1,146	556	590
4. Epilepsy	1,373	777	596
5. Alcohol use disorders	685	600	85
6. Alzheimer's and other dementias	450	197	253
7. Parkinson's disease	100	54	46
8. Multiple sclerosis	77	29	48
9. Drug use disorders	929	699	230
10. Post-traumatic stress disorder	224	61	162
11. Obsessive-compulsive disorder	619	254	365
12. Panic disorder	519	174	346
13. Insomnia (primary)	234	136	98
14. Migraine	329	92	237
15. Mental retardation, lead-caused	1,505	753	752
Other neuropsychiatric disorders	2,481	1,306	1,174
F. Sense organ diseases	8,939	3,980	4,959
1. Glaucoma	937	416	522

2. Cataracts	5,169	2,224	2,944
3. Vision disorders, age-related	920	397	522
4. Hearing loss, adult onset	1,912	942	970
Other sense organ disorders	2	1	1
G. Cardiovascular diseases	15,069	6,738	8,331
1. Rheumatic heart disease	479	165	314
2. Hypertensive heart disease	937	338	599
3. Ischemic heart disease	4,579	2,399	2,180
4. Cerebrovascular disease	5,125	2,077	3,048
5. Inflammatory heart diseases	945	490	455
Other cardiovascular diseases	3,004	1,268	1,736
H. Respiratory diseases	6,150	3,559	2,592
1. Chronic obstructive pulmonary disease	1,631	1,065	566
2. Asthma	1,925	1,074	851
Other respiratory diseases	2,595	1,420	1,175
I. Digestive diseases	7,226	3,836	3,390
1. Peptic ulcer disease	345	199	146
2. Cirrhosis of the liver	1,212	750	462
3. Appendicitis	44	25	19
Other digestive diseases	5,626	2,863	2,763
J. Genitourinary diseases	2,623	1,341	1,281
1. Nephritis and nephrosis	1,633	727	907
2. Benign prostatic hypertrophy	292	292	—
Other genitourinary system diseases	697	322	375
K. Skin diseases	956	482	473
L. Musculoskeletal diseases	2,171	975	1,196
1. Rheumatoid arthritis	252	94	158
2. Osteoarthritis	1,278	523	755
3. Gout	94	83	11
4. Low back pain	214	113	101
Other musculoskeletal disorders	333	162	172
M. Congenital anomalies	3,441	1,819	1,622
1. Abdominal wall defect	36	20	16
2. Anencephaly	47	23	24
3. Anorectal atresia	14	9	4
4. Cleft lip	12	6	6
5. Cleft palate	28	11	17
6. Esophageal atresia	2	1	1
7. Renal agenesis	2	2	0
8. Down syndrome	419	248	171
9. Congenital heart anomalies	1,651	849	802
10. Spina bifida	293	146	147
Other congenital anomalies	938	504	434
N. Oral conditions	720	347	373
1. Dental caries	496	248	248
2. <i>Periodontal disease</i>	23	11	12
3. Edentulism	181	80	101
Other oral diseases	21	8	13
III. Injuries	28,848	20,819	8,030
A. Unintentional injuries	18,876	12,491	6,385
1. Road traffic accidents	6,374	4,186	2,188
2. Poisonings	954	599	355
3. Falls	976	619	357
4. Fires	1,739	1,057	682
5. Drownings	1,720	1,294	426

6. Other unintentional injuries	7,112	4,736	2,377
B. Intentional injuries	9,972	8,328	1,645
1. Self-inflicted injuries	882	663	220
2. Violence	4,996	4,007	989
3. War	4,090	3,655	435
Other intentional injuries	3	2	1

Source: Lopez A, Mathers C, Ezzati M, Jamison DT, Murray CJLM, eds. *Global Burden of Disease and Risk Factors* (2006), available at <http://www.dcp2.org/pubs/GBD>.

Notes: — = an estimate of zero; the number zero in a cell indicates a non-zero estimate of less than 500.

DALYs refers to disability adjusted life-years.

a. Does not include liver cancer and cirrhosis DALYs resulting from chronic hepatitis virus infection.

b. This cause category includes “Causes arising in the perinatal period” as defined in the International Classification of Diseases, principally low birthweight prematurity, birth asphyxia, and birth trauma, and does not include all causes of DALYs(3,0) occurring in the perinatal period.

Table 2: Ranking of selected interventions based on WHO Choice Project by DALYs averted for WHO region AfrE

Intervention	Cost-Effectiveness Category	DALYs averted per 1 million persons per year
Treatment of new smear-positive TB cases under DOTS	Highly cost-effective	129,224
Peer education & treatment of STIs for sex workers	Highly cost-effective	55,635
Community support for exclusive breastfeeding & low birth weight baby care	Highly cost-effective	20,492
Case management of malaria with artemisinin-based combination	Highly cost-effective	16,989
Mass media campaign to promote safer sex	Highly cost-effective	13,103
Case management of neonatal & childhood pneumonia	Highly cost-effective	12,872
Oral rehydration therapy for diarrhea	Highly cost-effective	12,269
Prevention of mother-to-child transmission of HIV	Highly cost-effective	10,018
Insecticide-treated bednets	Highly cost-effective	9,387
Emergency neonatal care (low birth weight, infections, asphyxia & jaundice)	Highly cost-effective	9,106
Cataract extraction with lens implant	Highly cost-effective	8,970
Skilled birth attendance	Highly cost-effective	8,843
Tetanus toxoid	Highly cost-effective	8,840
Surgical treatment of trachoma	Highly cost-effective	8,051
Taxation rate of current + 50% on alcohol	Highly cost-effective	7,514
Treatment of smear-negative TB cases under DOTS	Highly cost-effective	7,514
Antiretroviral therapy (first-line drugs only)	Highly cost-effective	7,058
Blood pressure & cholesterol-lowering drugs & aspirin for 10-year CVD risk \geq 15%	Highly cost-effective	5,462
Treatment of sexually transmitted infections	Highly cost-effective	5,305
Vitamin A and Zinc supplementation	Highly cost-effective	5,280
Voluntary counseling and testing	Highly cost-effective	5,042
Indoor residual spraying	Highly cost-effective	4,570
Iron supplementation	Highly cost-effective	4,083
Antenatal care	Highly cost-effective	3,688
C-section & assisted vaginal delivery	Highly cost-effective	3,607
Fortification of food with Vitamin A and Zinc	Highly cost-effective	3,459
Reduction of unsafe injection use	Highly cost-effective	3,203
Taxation rate of 150% on tobacco	Highly cost-effective	2,986
Standardized second-line drug re-treatment of TB	Highly cost-effective	2,981
Antenatal steroids for pre-term births	Highly cost-effective	2,877
Biannual screening for breast cancer with appropriate treatment	Highly cost-effective	1,847
Treatment of severe pre-eclampsia/eclampsia	Highly cost-effective	1,525
Case management of depression with older drugs (TCAs)	Cost-effective	1,517
Reduction of injection use	Highly cost-effective	1,445

Older anti-epileptic drugs	Highly cost-effective	1,196
Improved indoor cooking stoves	Highly cost-effective	1,140
Facility-based care of maternal sepsis	Highly cost-effective	964
Brief advice by a physician for heavy alcohol users	Highly cost-effective	929
Advertising ban on alcohol	Highly cost-effective	784
Measles vaccination	Highly cost-effective	584
Antibiotics for pPROM	Highly cost-effective	545
Screening for retinopathy & photocoagulation	Highly cost-effective	531
Lithium for bipolar disorder with psychosocial treatment	Cost-effective	409
Typical anti-psychotic drugs for schizophrenia with psychosocial treatment	Cost-effective	353
School based education on safer sex	Highly cost-effective	350
Information dissemination for tobacco	Highly cost-effective	290
Facility-based care of severe post-partum haemorrhage	Highly cost-effective	250
Comprehensive advertising ban on tobacco	Highly cost-effective	243
Treatment of colorectal cancer	Cost-effective	241
Pop'n wide strategies to reduce blood pressure & cholesterol	Cost-effective	193
Clean indoor air laws	Cost-effective	133
Intensive glycemic control for diabetes	Cost-effective	83
Intermittent presumptive treatment of malaria during pregnancy	Highly cost-effective	22

Source: Choosing Interventions that are Cost Effective, World Health Organization, Geneva, available at <http://www.who.int/choice/en/>.

Notes: DALYs refers to disability adjusted life-years.

Cost-Effectiveness category cutoffs for highly cost-effective is less than or equal to GDP per capita per DALY averted. The cutoff for cost-effective is between 1&3 times the GDP per capita per DALY averted.

GBD region AfrE consists of: Botswana, Burundi, Central African Republic, Democratic Republic of Congo, Republic of Congo, Cote d'Ivoire, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Tanzania, Uganda, Zambia, & Zimbabwe.