INTERNATIONAL UNION FOR THE SCIENTIFIC STUDY OF POPULATION

Monitoring demographic indicators for the post 2015 Sustainable Development Goals (SDGs)

A review of proposed approaches and opportunities

Prepared by Stephane Helleringer, Johns Hopkins University

4/1/2015

This report reviews the indicators proposed by the Sustainable Development Solutions Network (SDSN) for the post-2015 SDG monitoring period that require access to population data or refer to demographic processes. We make recommendations to strengthen the proposed monitoring framework. The report was conducted as part of the IUSSP's activities related to the post-2015 data revolution with funding from UNFPA.

Executive Summary of recommendations

Based on a review of available data sources and estimation strategies, we suggest that low and middle-income countries where complete vital registration systems do not yet exist, or cannot be established in a short time, should adopt a tiered SDG monitoring framework. This framework combines

- High-quality decennial censuses
- Annual surveys of the proximate determinants of fertility/mortality and
- Periodic large surveys of fertility/mortality with verbal autopsies (every 3-5 years).

This proposal thus differs from current calls for annual reporting on all SDG indicators. Such high frequency will not be possible for key mortality and fertility indicators in LMICs with limited vital registration systems. This is so because, on the one hand, new initiatives to produce yearly estimates of these key demographic processes (e.g., model-based strategies, community-based key informants) are indeed affected by large biases. On the other hand, prohibitively large surveys would be required to monitor mortality and fertility on an annual basis. This is primarily the case because births and deaths remain rare events (approximately 10-40 per 1,000 population), which exhibit limited year-to-year variations.

For such a tiered monitoring framework to yield unbiased assessments of SDG progress however, an important program of methodological research should be launched. The primary focus of this program of research should be on increasing the quality of survey data on fertility and mortality. A second emphasis of this program of research should aim at improving estimates of population sizes in intercensal years and for population sub-groups. This latter aim will require improving the quantity and quality of data available on migration both within countries and across national borders.

We also i) suggest the addition of two indicators of the proximate determinants of fertility; ii) recommend several modifications to the definitions of indicators proposed by the SDSN, and iii) suggest additional data collection initiatives.

Finally, we make recommendations to enable the disaggregation of trends in SDG indicators across population sub-groups. This central ambition of the SDGs will indeed require:

- Planning surveys and sample size calculations, so that differences in trends between sub-groups ("difference-in-differences") can be detected over time,
- Developing survey instruments that eliminate differential reporting across population sub-groups
- Adopting a simple analytical framework based on standardization and decomposition, which permit identifying situations of convergence/divergence between population sub-groups.

LIST OF ACRONYMS AND ABBREVIATIONS

ACASI	Audio Computer-Assisted Self-Interviewing
BH	Birth History
cDHS	Continuous Demographic and Health Survey
DHS	Demographic and Health Survey
ELSA	English Longitudinal Study of Ageing
GDP	Gross Domestic Product
HDSS	Health and Demographic Surveillance System
HLE	Healthy Life Expectancy
HRS	Health and Retirement Study
ICD	International Classification of Diseases
IHME	Institute for Health Metrics and Evaluation
IUD	Intrauterine Contraceptive Device
IUSSP	International Union for Scientific Study of Population
LMICs	Low and Middle Income Countries
M&E	Monitoring and Evaluation
MCOD	Multiple Causes of Death
MDG	Millennium Development Goal
MICS	Multiple Indicators Cluster Survey
MMR	Maternal Mortality Ratio
NCD	Non-Communicable Disease
ODK	Open Data Kit
OECD	Organization for Economic Development and Cooperation
OWG	Open Working Group
SAGE	WHO Study of Global Ageing and Adult Health
SDG	Sustainable Development Goals
SDSN	Sustainable Development Solutions Network

SHARE	Survey of Health, Ageing and Retirement in Europe
SPD	Sentinel Panel of Districts
SSH	Siblings' Survival History
TB	Tuberculosis
TFR	Total Fertility rate
UNFPA	United Nations Population Fund
UNICEF	United Nations Children Fund
UNOCHA	UN Office for the Coordination of Humanitarian Affairs
UNPD	United Nations Population Division
UN MMEIG	UN Maternal Mortality Estimation Inter-agency Group
UNSC	UN Statistical Commission
VA	Verbal Autopsy
WHO	World Health Organization
WHS	World Health Survey

TABLE OF CONTENTS

Section I. Introduction

The sustainable development goals (SDGs) aim to extend and improve the approach of the Millennium Development Goals beyond 2015. One area where significant improvements are needed is the area of monitoring and evaluation (M&E). MDGs presented significant M&E challenges.

- They were adopted in 2000 but set targets relative to the situation of the world 1990: this generated a need to ascertain most MDG baseline indicators retrospectively, using patchy information and data sources. The baseline level of MDG indicators in 1990 is thus often heatedly debated, with consequences for assessments of the progress towards MDG targets. For example, even though the UN Maternal Mortality inter-agency estimation group (UN MMEIG) and the Institute for Health Metrics and Evaluation (IHME) agree on the global number of maternal deaths in 2012-2013, the UN MMEIG estimates that there were 545,000 maternal deaths in 1990 (Zureick-Brown et al., 2013, Wilmoth et al., 2010), whereas IHME suggests that there were 376,000 such deaths in 1990 (Kassebaum et al., 2014). The UN MMEIG thus estimates that maternal deaths have declined much faster than IHME estimates.
- Since 2000, MDG indicators are often only partially reported: only a small number of countries report all required indicators. In addition, when data are available, they are also often reported with significant delays, possibly several years after completion of data collection (Rugg et al., 2009, United-Nations, 2014).
- Some indicators (e.g., maternal mortality ratios) of progress towards the achievement of the MDGs also cannot be measured accurately, e.g., due to limitation of input datasets or because they constitute rare events, which require very large data collection undertakings (El Arifeen et al., 2014, Hill et al., 2006). As a result, some MDG indicators may require a number of proxy measures.

Ultimately, it is difficult to decipher how different programs and inputs have contributed to MDG progress, and the “return on MDG investments” is frequently unknown. Such deficiencies in M&E have likely hampered MDG progress: effective interventions and scalable programs are not identified rapidly enough, possibly effective schemes are abandoned and scarce resources are not allocated towards activities that would produce

the maximum “bang for each buck”. For these reasons, it is important that a robust monitoring framework for the SDGs be put in place as early as possible.

The development of the SDGs and the associated monitoring framework is a complex, multi-level process, which was initiated several years ago. Heads of State will formally agree upon the SDGs in September 2015. A set of indicative SDG indicators should be adopted around the same time, so that (1) the UN statistical commission can ultimately adopt the monitoring framework for the SDGs early in 2016, and (2) baselines for each of the indicators can be established before or near the start of the SDG period. Various consultations and reports coordinated by the Open Working Group on the SDGs will play a key role in framing the deliberations of the UN statistical commission (UNSC). These include in particular several reports, briefs and evidence papers developed by the Sustainable Development Solutions Network (SDSN). In this report, we discuss such existing proposals for SDG indicators and monitoring frameworks.

(a) The key functions of SDG indicators

The overarching goals of SDG indicators should be two-fold: to serve as a “management tool” and to serve as a “report card” for development programs (SDSN, 2015). In this report, we evaluate the ability of a selected subset of recently proposed SDG indicators focused on demographic processes to accomplish these two functions.

As a management tool, SDG indicators should provide key information on the ongoing implementation of development programs. Such information is required to guide scale-up programs, intensify activities as required, and adopt possible course corrections. To fulfill this role, SDG indicators must be updated frequently, in order to capture emerging challenges, bottlenecks and identify areas where programs are weak. Currently, there are numerous calls for yearly reporting on these indicators with suggestions that some indicators could even be updated more frequently (e.g., similar to reports on economic growth in OECD countries, or employment reports in more developed economies). We will assess the possibility of producing annual updates of SDG indicators related to demographic dynamics.

As report cards, on the other hand, SDG indicators are expected to reflect the performance of particular development programs so as to ensure accountability of various actors involved in these programs (e.g., governments, NGOs, international organizations). To fulfill this role, it is thus crucial that SDG indicators are not confounded by simultaneous trends in other variables and development processes. For example, consider a disease is strongly associated with poverty. If development programs are successful at significantly reducing poverty in a given country, then we should expect the prevalence of that disease to decline independently of any disease-specific or health systems interventions. An indicator that amalgamates the effects of poverty on this disease with effects of health interventions would not be adequate as a "report card". Instead, we need an indicator that isolates the effects of health interventions from the effects of other concomitant factors. In this report, we will assess the capacity of various SDG indicators to control for such confounders.

(b) Leaving no one behind: monitoring SDG indicators in sub-groups

Reducing inequalities is a major focus of the SDGs (Sachs, 2012). One goal (Goal #10) thus entails "reducing inequality within and among countries", whereas several other goals place inclusiveness and equity at the center of the SDGs (e.g., "promote inclusive industrialization", "make cities and human settlements inclusive"). This has important implications for the monitoring and evaluation of SDG programs: ideally, SDG indicators should be monitored not only at the national level, but also within pre-defined population sub-groups, including for example by sex, age, residence (rural vs. urban) or wealth status.

This additional disaggregation requirement for an SDG monitoring framework presents challenges however. On the one hand, it requires ensuring that the quality and completeness of key data sources (e.g., vital registration, census) does not differ between population sub-groups. For example, if death registration is higher among men than among women, then gender differences in mortality may appear more/less pronounced than they really are. On the other hand, disaggregation will also require significantly larger investments in survey data collection. Obtaining precise measurements of SDG indicators in multiple sub-groups will indeed necessitate

increasing sample sizes. For example, the sample size of Demographic and Health Surveys (DHS) has increased by orders of magnitude between earlier waves when the main objective was to obtain national-level estimates, and the most recent waves, which also often aimed to produce precise estimates by regions or districts. Finally, the additional disaggregation requirement will also necessitate the development of simple analytical strategies that permit rapidly identifying situations of growing vs. declining inequality. Current tools available to measure inequalities (e.g., Gini coefficients, dissimilarity indexes) may not be adapted to future SDG indicators and may occasionally be difficult to interpret. We thus outline, discuss and propose strategies to address the additional issues raised by the necessity to disaggregate SDG indicators between various population sub-groups.

(c) Scope of the report

In discussing SDG indicators, we use the existing proposals emanating from the SDSN (SDSN, 2015). This the most comprehensive list, which will serve as the basis of future discussions and deliberations of the UNSC. At the moment, 100 SDG indicators have been proposed to measure progress towards 18 goals (SDSN, 2015). Several indicators are inherited from the MDG monitoring framework, in large part to ensure continuity of time-series and to permit long-term assessment of progress towards targets such as the eradication of absolute poverty or the fight against diseases. Other indicators have been developed *de novo*, whereas others still remain in development. Our assessment focuses on the subset of the proposed SDG indicators that are demographic in nature.

Demographic indicators are indicators that reflect fundamental aspects of population dynamics such as fertility, mortality and/or migration. Indicators of this kind figure prominently among the list of proposed SDG indicators, as was the case during the MDG period. Most of the demographic indicators included in the provisional list of SDG indicators fall under goal #3 (“ensure healthy lives and promote well-being for all at all ages”) and goal #5 (“achieve gender equality and empower all women and girls”), even though another demographic indicator is also listed under goal #16 (“promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels”). The full list of

demographic SDG indicators is shown in table 1 below, along with a description of each indicator.

By contrast, we employ the term “population-based indicator” to refer to other SDG indicators that require an accurate count of the population-at-risk for their calculation. Such indicators include, for example, the proportion of the population in extreme poverty, tertiary enrollment rates for women and men or the share of the population with access to modern cooking solutions. Our ability to consistently and frequently measure these indicators may be affected by limited information on the size and composition of populations. Population-based indicators are not the focus of our report, but we will discuss some of the measurement issues that are common to demographic and population-based indicators. These issues stem in particular from inaccuracies on population size and composition contained in census data, and the difficulty and uncertainty inherent in estimates of population size and composition in intercensal years (e.g., because of migration).

Other indicators included in the proposed list of SDG indicators consist of indicators that traditionally fall outside of the realm of demography and/or do not require estimates of population size for computation. This is the case, for example, of indicators related to goal #13 (“Take urgent action to combat climate change and its impact”), goal #14 (“Conserve and sustainably use the oceans, seas and marine resources for sustainable development”) and goal #15 (“Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss”). This does not mean however that demographic trends do not impact these indicators. For example, population dynamics play a key role in determining the levels of CO2 emissions (Zagheni, 2011).

Goal	Indicator	Indicator description	Standard data source(s)	Responsible agency
1	7	Total Fertility Rate	Census + vital statistics + survey data	UNPD, UNFPA
2	11	Percentage of children less than 6 months who are fed breast milk alone	Survey data	WHO, UNICEF
3	17	Maternal Mortality Ratio and Rate	Census + vital statistics and/or hospital data	WHO, UNPD, UNICEF, WORLD BANK
3	18	Neonatal, Infant and Under-5 Mortality Rates	Census + vital statistics	WHO, UNPD, UNICEF
3	20	HIV Mortality	Census + vital statistics and/or hospital data	WHO, UNAIDS
3	21	TB Mortality	Census + vital statistics and/or hospital data	WHO
3	22	Malaria Mortality	Census + vital statistics and/or hospital data	WHO
3	23	Probability of dying between 30 and 70 from any cardiovascular disease, cancer, Diabetes or chronic respiratory disease	Census + vital statistics and/or hospital data	WHO
3	25	Road traffic deaths per 100,000 population	Census + vital statistics and/or police/hospital reports	WHO
3	29	Contraceptive Prevalence Rate	Survey data	UNPD, UNFPA
5	40	% of women aged 20-24 years old who were married or in a union before age 18	Census + vital statistics	UNICEF
5	44	Met demand for family planning	Survey data	UNPD, UNFPA
16	88	Violent injuries and deaths per 100,000 population	Census + vital statistics and/or police/hospital reports	WHO, UNOCHA

Table 1: List of Demographic Indicators included in the report; *notes: standard data sources refer to the sources, which would be used in contexts where vital registration systems and health information systems have high coverage and accuracy. Responsible agency denotes the UN body, which is currently responsible for gathering data and reporting on proposed indicator.*

Several of the other indicators included in the proposed list of SDG indicators also lend themselves to demographic analyses, e.g., using life tables and/or decomposition techniques. For example, trends in the proportion of women in parliament (SDG indicator

#43) could be analyzed using such techniques. In that case, we would consider that the members of parliament constitute the population of interest, which is distributed by sex, age and/or number of times elected. Then we could decompose changes in the number of women in parliament into the relative contributions of a) changes in the proportion of women among first-time parliament members and b) gender differences in likelihood of re-election. This knowledge would help orient interventions and actions supporting the participation of women in the political sphere. Similar analyses have been conducted, for example, for the Supreme Court of the US (Stolzenberg, 2011, Stolzenberg and Lindgren, 2010)

(d) Approach to assessment of demographic SDG indicators

We assess the feasibility of measuring the proposed demographic SDG indicators. These indicators are designed for universal use, i.e., they can be measured and should be informative for every country, and they should be comparable across countries. There are however profound inequalities in data availability worldwide (AbouZahr et al., 2007, Setel et al., 2007, Jha, 2014, Hill et al., 2007, Jha et al., 2007). In our assessment, we will thus differentiate between high-income countries (HIC) and low and middle-income countries (LMIC). HICs typically have information systems that permit the routine monitoring of all demographic and population-related proposed SDG indicators. The complete vital registration of births, marriages and deaths constitutes the cornerstone of these information systems. Accurate information on population size and composition is either provided by regular high-quality censuses, or obtained from the triangulation of administrative databases. In LMICs, on the other hand, the state of information systems is significantly more heterogeneous. In particular, vital registration systems are often too incomplete to permit the monitoring of the proposed demographic and population-related SDG indicators. In LMIC settings, the adoption of the SDGs must thus be matched by a "data revolution", i.e., a comprehensive process aimed at improving the quantity and quality of data available for development programs (Atun, 2014, Mitra, 2013).

Our report will be primarily focused on the situation of demographic measurement in LMICs, and particularly data availability in sub-Saharan countries. Vital registration systems in sub-Saharan countries indeed often have very low coverage rates. Basic

demographic rates are thus produced using unconventional sources or techniques. In order to ensure that SDG indicators can be compared across countries, significant investments in data collection will need to be targeted at sub-Saharan countries where data availability and data quality are the lowest. Ideally, these investments would permit increasing the completeness of vital registration systems to levels where vital registration data can be used to calculate demographic indicators. In this report however, we do not make the assumption that such investments will materialize in the short to medium term, despite the fact that this must be our long-term goal. Instead, we discuss primarily interim strategies that can be used to supplement defective vital registration systems during the 2015-2030 period.

We build on expertise in data collection and analysis accumulated by demographers working in HICs and LMICs. In particular, we mobilize models and techniques designed to produce demographic estimates from imperfect data sources (UN, 1983). These tools play a key role in measurement, monitoring and evaluation of demographic trends in LMICs. Most of the expertise of demographers in handling imperfect data was initially summarized in several manuals of the United Nations (e.g., UN, 1983). It was subsequently updated (Moultrie et al., 2013) by a working group of the International Union of the Scientific Study of Population (IUSSP). This expertise then informs reviews of available data sources and discussions of proposed innovative data collection approaches (e.g., big data).

The data reviews we conduct are illustrative, rather than exhaustive. They are aimed at highlighting typical challenges that will need to be overcome in order to enable monitoring of SDG indicators. We will occasionally explore specific examples of measurement issues and data collection approaches in more detail. When we do so, it is with the idea that the case investigated presents broader lessons that are applicable in a wide array of settings.

The discussions and recommendations contained in the report are intended to inform the debate about an SDG monitoring framework and a list of SDG indicators. As such, they are not intended for a particular organization or body, even though they may be particularly of particular concern for the entities in charge of monitoring a specific

indicator.[1] Some of the recommendations may also be focused on issues of data collection and thus are more relevant for groups actively engaged in data collection, e.g., groups running the demographic and health surveys or the Multiple indicator Cluster Surveys. Other recommendations may on the other hand be more conceptual and related to the definition of specific indicators. These recommendations may be more relevant for groups engaged in making proposals of indicators to the UNSC (e.g., the SDSN). Finally, another set of recommendations and discussion will focus on analytical issues, e.g., as they relate to the disaggregation of SDG indicators between population sub-groups. These recommendations will be more relevant for various institutes and reference groups engaged in producing estimates of demographic indicators (e.g., UN MMEIG, UNPD, IHME).

(e) Organization of the report

The report is organized as follows. In section 2, we survey the various data sources currently available on demographic indicators. We highlight differences in data availability between HICs and LMICs, and we also investigate differences among LMICs. In section 3, we review current approaches to producing estimates of the proposed demographic SDG indicators from these data. We emphasize existing global initiatives (e.g., IHME's Global Burden of Disease Study), which seek to produce comparable estimates of these indicators despite large differences in data availability and quality. We argue that the model-based approaches these initiatives have adopted may lead to significant biases in estimates of (trends in) SDG indicators. As a result, they are not well suited to serve as "management tools" and "report cards" for programs tackling the SDGs. In section 4, we assess the ability of the proposed demographic SDG indicators to serve as "report cards" for development and health programs. In doing so, we suggest several modifications to the definition of proposed SDG indicators, which are required to avoid conceptual errors and/or confounding from other concomitant processes. We also emphasize several analytical strategies and data collection instruments that may help obtain informative measurements of SDG indicators. In section 5, we investigate the possibility of accelerating data collection on the proposed demographic SDG indicators,

[1] This specific organizations and UN bodies are listed in the rightmost column of table 1.

in order to obtain robust annual time-series of indicators. In particular, we review the role of new technologies constitute in facilitating this acceleration. We suggest however that, since most demographic events remain rare events (e.g., on the order of 10-30 per 1,000 person-years), even yearly data collection will seldom permit detecting year-to-year changes in some of the most important demographic SDG indicators unless prohibitively large surveys are conducted. We thus propose a two-tiered data collection system for SDG monitoring in LMICs. Finally, in section 6, we tackle the question of disaggregation, i.e., monitoring SDG indicators separately in population sub-groups.

Section II. Overview of data sources on proposed demographic SDG indicators

KEY POINTS

- In high-income countries, the administrative databases, vital registration systems and survey programs required to monitor the proposed demographic SDG indicators are in place and functioning.
- In low and middle-income countries, these data sources are often unavailable or incomplete (infrequent censuses, limited vital registration, inaccurate cause of death certification)
- Instead, the data required to monitor the proposed demographic SDG indicators are often obtained only from retrospective surveys
- But significant data gaps remain since these surveys are conducted infrequently and have not yet been fielded in a significant number of countries.

Sub-Section 2.01 Mortality Indicators

Nine of the currently proposed demographic SDG indicators are mortality indicators. These include cause-specific death rates (e.g., malaria death rate) as well as life table quantities (e.g., NCD-specific ${}_{40}q_{30}$). In this sub-section, we review and synthetize current data sources available for the measurement of these indicators.

(a) High-income countries:

In HICs, the measurement of mortality indicators is relatively straightforward. The required data on events and person-years are available from a combination of census data, vital registration data, as well as administrative data on traffic-related or violent deaths. Data on specific diseases from hospital registries may also be used in calculating cause-specific death rates (e.g., cancer). Difficulties may arise because of

occasional age misreporting among the oldest-old however (e.g., Preston et al., 1996), as well as misclassifications of causes of death in vital registration data (Kassebaum et al., 2014, Kao et al., 1997). Maternal mortality, for example, can be under-reported in vital registration systems in HICs: maternal deaths are often classified as deaths from other causes. Procedures exist however, which permit adjusting raw vital registration data for such misclassifications and errors.

The calculation of healthy life expectancy (HLE) at birth, on the other hand, is more contested. HLE refers to the number of years one can expect to live in good health. It requires data on mortality by age from vital statistics, but it also requires data on health states (e.g., health vs not healthy) and transitions between these states. These data are usually collected during cross-sectional or longitudinal surveys. Examples of such surveys include the health and retirement study (HRS) in the US, the English Longitudinal Study on Aging (ELSA) or the Survey of health, aging and retirement in Europe (SHARE). HLE is then calculated using statistical models that combine the various required data sources. In recent years, the Sullivan Method for calculating HLE has been contested and several new approaches have been proposed.

(b) Low and middle-income countries:

In LMICs, the estimation of mortality indicators presents significantly more challenges, primarily because of data limitations. Even though the availability and quality of different data sources required for mortality estimates vary significantly across LMICs, most LMICs often do not derive mortality estimates from vital registration data. Instead, they resort to alternative data sources that are easier to collect, e.g., survey datasets. We review the availability of each type of data across LMICs.

(i) Census data

The UN recommends that censuses are conducted every 10 years or so in every LMICs. They provide the basic population counts to be incorporated in the calculation of demographic rates, often by age, sex, educational level and/or poverty status. Censuses conducted in LMICs have also increasingly included retrospective questions about all the deaths that have occurred in a household over the 12 months before the census (Hill et al., 2007: Hill et al., 2006: Whiting et al., 2006). These questions permit calculating various mortality rates, including for example maternal mortality rates.

The frequency and quality of census implementation has greatly increased in LMICs since 2000. Some countries affected by civil conflict recently conducted their first census in decades (e.g., Angola conducted its first census in 44 years in 2013). The most recent censuses have also incorporated new technologies of data collection, thus making data available much more rapidly. For example, the results of the most recent Senegal Census were available only a few months after the end of fieldwork due to data collection on PDAs.

Significant challenges remain however: in some countries, censuses are still infrequent, whereas in others, they are contested for political and/or technical reasons (e.g., Nigeria). Post-census enumeration surveys are also not systematically conducted; as a result estimates of census coverage are often imprecise and undercounts may be undetected. The denominators of most rates of interest thus remain difficult to obtain and are affected by significant uncertainty. Finally, there are also important concerns about the reliability of estimates of mortality rates obtained from retrospective questions on household deaths. Some deaths may be omitted during census interviews (reporting errors), whereas other cannot be counted because the households where they occurred may have dissolved prior to the census. On the other hand, census data on mortality may be affected by double counting if some deaths are reported in two households (e.g., the deaths of a polygamous husband being reported by two wives living in different households). Census data on mortality may thus be affected by complex biases.

(ii) Vital registration

Ideally, vital registration, i.e., the continuous recording of births, deaths and marriages, a) provides the counts of events that constitute the numerator of demographic rates, b) permits updating estimating the numbers of person-years lived in a population between two censuses, and c) provides information on the causes of recorded deaths. In LMICs however, vital registration systems often do not play any of these 3 roles. They have limited coverage, are not timely reported and provide inaccurate data on causes of death. The extent of under-registration of deaths in LMICs is shown in Figure 1.

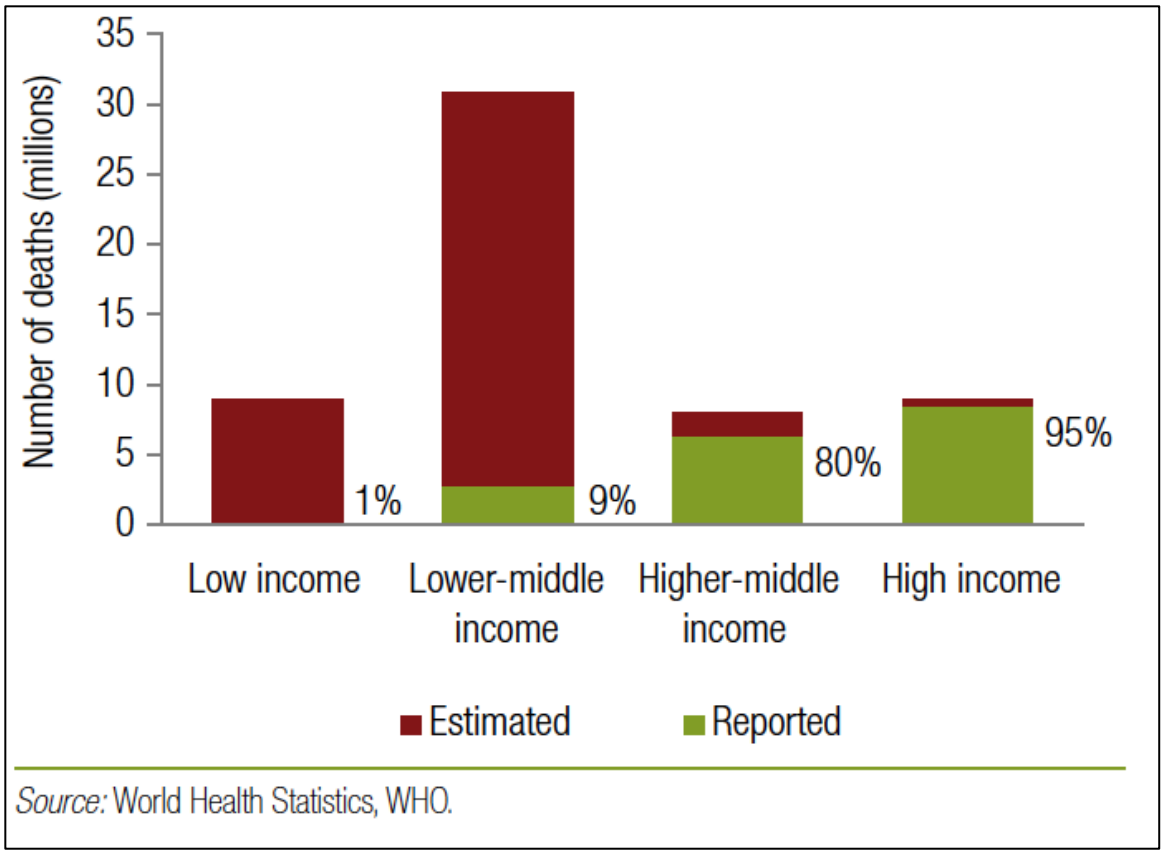

Figure 1: Sources of information about deaths across countries, by income level (source: World Bank / WHO, in Global Civil Registration and Vital Statistics Scaling Up Investment Plan, 2015-2024**)**

The situation of **death registration** (Bank/WHO, 2014) varies significantly across and within LMICs (see Figure 2). It is generally strongest in Latin American countries (Danel and Bortman, 2009), whereas countries in sub-Saharan Africa and Southeast Asia have very low coverage rates (Jha, 2014, AbouZahr et al., 2007, Mahapatra et al., 2007). Even within each region, some countries perform better than others: for example, in Latin America, the coverage of vital registration is > 90% in countries such as Argentina, Chile or Costa Rica, whereas coverage rates in countries such as Bolivia, Honduras or Nicaragua are significantly lower (e.g., between 25 and 70%). In sub-Saharan Africa, death registration is high in a few Island States (e.g., Mauritius), and it is also > 80% in South Africa. In West African countries, on the other hand, death registration is often less than 20%.

Some (large) countries (e.g., India) have also adopted sample vital registration systems, which permit producing estimates of demographic rates even if only a small percentage of the country's population is actually covered by vital registration.

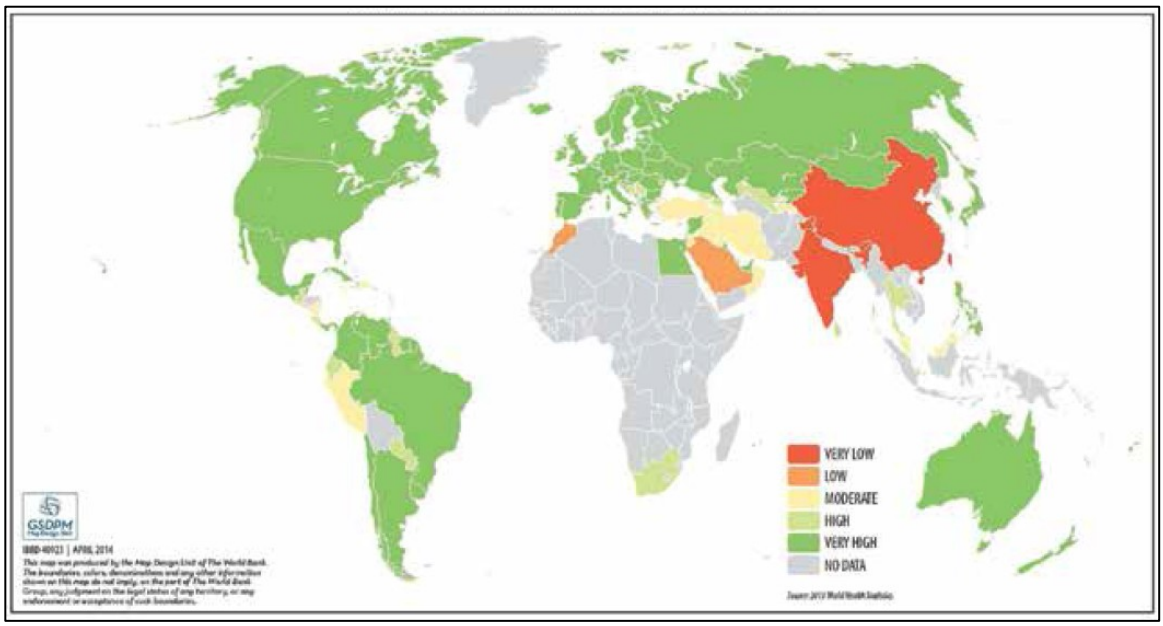

Figure 2: Coverage of death registration by country in 2012; source WHO statistics in World Bank / WHO Global Civil Registration and Vital Statistics Scaling Up Investment Plan, 2015-2024

Within LMICs, there are often large differentials in death registration between urban and rural areas, and across socioeconomic groups. In urban areas indeed, a death certificate is often required for burials, whereas this is not the case in rural areas. Death registration is also often highest among those employed in the formal sector in urban areas, since benefit claims also require obtaining a death certificate. Finally, the completeness of death registration may vary by age at death: adult deaths are frequently much more completely registered than deaths among children.

The data on causes of death that can be obtained from vital registration in LMICs is also often questionable (see Figure 3). Causes of death are frequently not reported at all. In other cases, so-called "garbage codes" or ill-defined causes of death are over-represented among registered causes of death, compared to data from more accurate and complete vital registration systems. This is the case because medical personnel are

often not able to provide information about the underlying cause of death, or at least about the process leading up to death (e.g., home deaths). Finally, some causes of death may be systematically under-recorded by vital statistics officers. This is the case in particular of HIV-related deaths. In South Africa, for example, HIV-related deaths are frequently misclassified as deaths from other, possibly unrelated causes (Dorrington et al., 2000: Dorrington et al., 2002: Kerber et al., 2013)

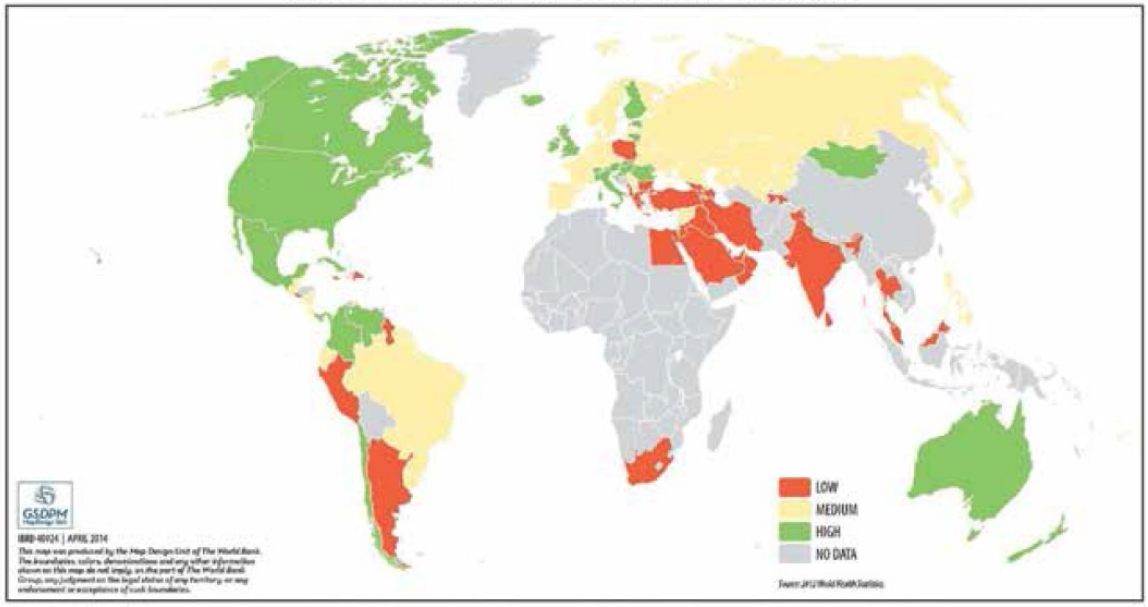

Figure 3: Quality of cause of death statistics by country in 2012, source WHO statistics in World Bank / WHO Global Civil Registration and Vital Statistics Scaling Up Investment Plan, 2015-2024

(iii) Administrative data

Several other data sources can provide information required to estimate specific mortality rates in LMICs. This includes primarily police reports of accidents and violent deaths, as well as hospital records of deaths.

(iv) Supplementary data sources

In the absence of complete vital registration data, mortality data are collected from a number of other sources in LMICs. Demographers and epidemiologists then use these data to produce estimates of key mortality rates. We highlight several data sources that constitute the key inputs for estimating mortality in LMICs.

- Retrospective mortality surveys: mortality data are often collected in LMICs during household surveys by asking respondents to provide information on the survival of their close relatives (Gakidou et al., 2004, Hill et al., 2005). Specifically, respondents are asked to report the full list of a subset of their relatives (e.g., children, siblings, spouse); then they are asked to report whether each of the nominated relative is still alive at the time of the survey; if a relative is deceased, they are asked how old s/he was when s/he died, how long ago s/he died; in some instances, respondents are also asked to report the circumstances of their relative's death so that the cause of death can be ascertained. Such methods can produce estimates of the MMR, as well as the childhood mortality rates (neonatal mortality, infant mortality and under-5 mortality). They are also often used to calculate probabilities of survival at adult ages (e.g., ${}_{45}q_{15}$). Birth histories (BH) typically serve to estimate childhood mortality rates, whereas siblings' survival histories (SSH) serve to estimate adult mortality rates. Several large-scale survey initiatives now systematically include the collection of such retrospective mortality data: the Demographic and Health Surveys (DHS) and the World Health Surveys (WHS) collect both BH and SSH; the Multiple Indicator Cluster Surveys (MICS), on the other hand, routinely collect BH but only seldom collects SSH (one exception is the recent 2014 Guinea-Bissau MICS). The coverage of DHS and MICS surveys is large, with most LMICs having conducted one or more of each survey in the past 20 years (see Figure 4 below). In some countries however, the most recent DHS or MICS survey may have been conducted several years ago: for example, in Angola, no survey has recently been conducted that included the SSH module, so that the recent level of adult/maternal mortality is difficult to ascertain directly. Survey-based data on mortality also present several limitations including high sample size requirements and limited statistical power, possible reporting errors and sampling selection biases. We develop these limitations further below.

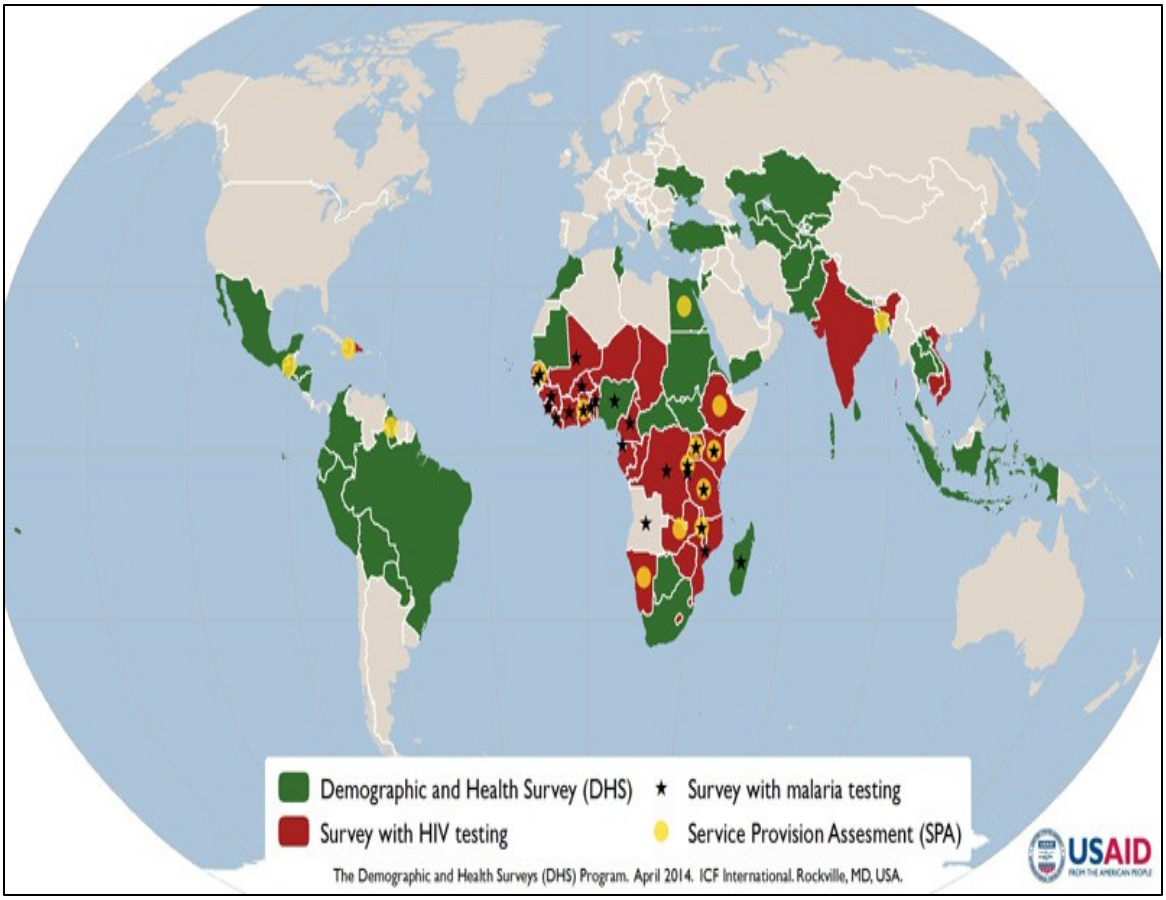

Figure 4: Availability of DHS datasets, worldwide, source: ICF-International Website, accessed on Dec 9th 2014)

- Longitudinal mortality studies: In parallel to the establishment of retrospective mortality surveys, longitudinal studies have been developed in several LMICs, often with the specific goal to measure mortality rates. A significant number of these longitudinal studies are health and demographic surveillance systems (HDSS), in which the dynamics of a small population are monitored through repeated household visits. These HDSS form the INDEPTH Network (Jha, 2014, Bangha et al.. 2010, Sankoh et al., 2006,Ngom et al., 2001). Longitudinal studies in LMICs permit obtaining direct estimates of mortality rates, but rarely provide nationally representative data. Exceptions include, for example, the Swaziland HIV Incidence Monitoring Survey (Bicego et al., 2013). Significant Hawthorne effects may also affect HDSS and other longitudinal studies, i.e., individuals modify their behaviors because they are under repeated observation (Zwane et al., 2011).

- Verbal autopsies: in the absence of certification of causes of death through vital registration systems, verbal autopsies (VA) have been used to measure the proportion of deaths attributable to specific causes (Chandramohan et al., 1998, Jha, 2014, Aleksandrowicz et al., 2014, Misganaw et al., 2012, Midhet, 2008, Murray et al., 2007, Setel et al., 2006, Quigley et al., 2000, Mobley et al., 1996). VAs require interviews with a close relative of the deceased, to inquire about symptoms and circumstances of the death. When available, VAs may be supplemented by medical records and other diagnostic procedures. On the basis of this information, one or more physicians then attribute a cause of death using ICD codes. Recently, methodological work has explored the feasibility of assigning causes of death through the use of statistical models (Byass et al., 2013, Byass et al., 2012, Ramroth et al., 2012, Vergnano et al., 2011, Fottrell et al., 2011, Tensou et al., 2010). VAs are systematically collected by HDSS. They have also been incorporated into some DHS or MICS surveys, and there are now calls to integrate VAs into vital registration systems of LMICs (Sankoh and Byass, 2014).

Sub-Section 2.02 Fertility and marriage indicators

The measurement of fertility and marriage indicators relies on some of the same data sources used for the monitoring of mortality indicators. Surveys however play a bigger role in fertility measurement since some indicators require reports of behaviors (e.g., contraception) or intentions (e.g., wanted fertility). Such data points are not readily collected in censuses and during vital registration. In addition, since births are repeated events, it is often easier to obtain precise estimates of fertility levels and trends using much smaller samples than those needed for mortality measurement.

(a) High-income countries:

In HICs, as in the case of mortality indicators, estimates of fertility/marriage rates again use census data and/or vital registration records. Measurement debates about key fertility rates have recently been focused on issues of “demographic translation”, i.e., approaches to inferring the fertility experience of cohorts from period data (Parrado, 2011). This debate was prompted primarily by increasing postponement of births in

HICs. This led to concerns that the period TFR may underestimate the true level of fertility in a population. Various adjustments have been proposed but none have been accepted as best practice. National survey programs supplement censuses and vital registration for the measurement of contraceptive prevalence.

(b) Low and middle-income countries:

Similar to mortality indicators, vital registration systems often do not permit measuring fertility and marriage rates in most LMICs without supplementary datasets and/or statistical adjustments.

In LMICs, the coverage of birth registration is generally higher than the coverage of death registration. This is often the case because birth certificates are required to enroll in school. Birth registration however seldom reaches levels that permit obtaining accurate estimates of fertility rates (see Figure 5 below). For example, in Latin America, in countries like Paraguay or Peru, only ≈ 50% of births are registered. In sub-Saharan countries, birth registration rates are less than 30% in a number of countries. In addition, children are often not registered immediately after birth but rather after significant delays. This may lead to errors in recorded ages. In LMICs where multiple censuses are available, fertility estimates can also be obtained by examining cohort parity increments (UN, 1983, Preston et al., 2001).

In these contexts of limited data availability, retrospective surveys often constitute the key source of information on fertility in LMICs. Birth histories included in such surveys permit estimating age-specific fertility rates, as well as a host of other fertility indicators not currently included as SDG indicators (e.g., interval length). Birth histories were first included in questionnaires of the World Fertility Survey, and now constitute the core of most DHS and MICS surveys (Shah et al., 1986, Chidambaram and Pullum, 1981). They provide a list of children born in the past 5 years before the survey, and for which additional questions on health, vaccinations and education will be asked. Standard calculations of the total fertility rate are thus available from DHS and/or MICS reports.

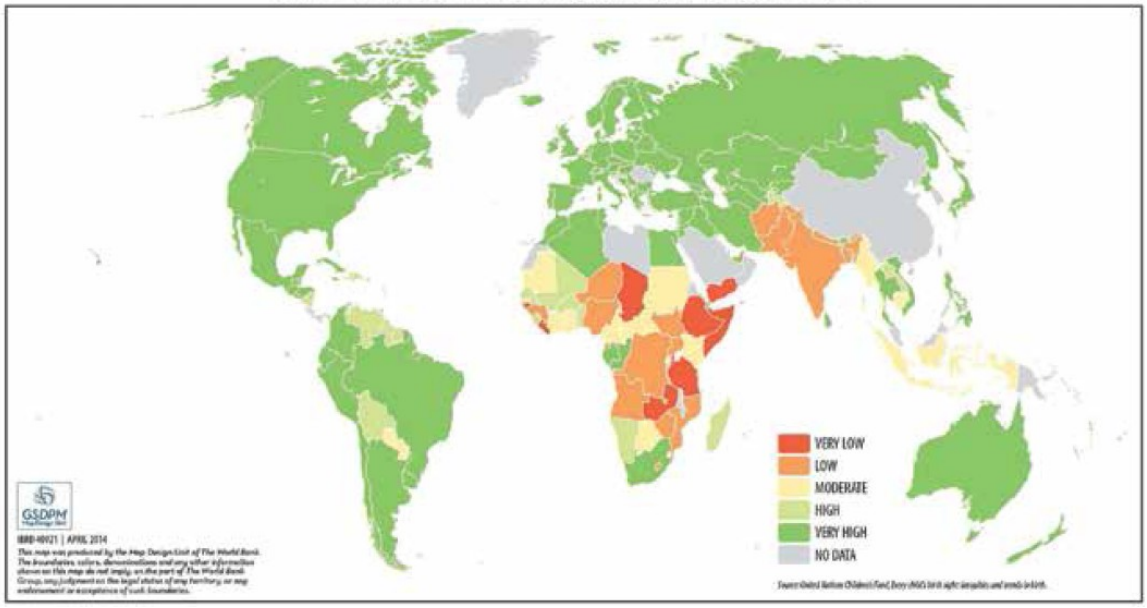

Figure 5: Coverage of birth registration by country, source: WHO statistics in World Bank / WHO Global Civil Registration and Vital Statistics Scaling Up Investment Plan, 2015-2024

Surveys also allow measuring the prevalence of contraceptive use, classifying recent births as wanted or unwanted, and measuring the extent of unmet need for family planning in a population (Pearson and Becker, 2014, McCoy et al., 2014, Asnake et al., 2013, Speizer et al., 2013, Alkema et al., 2013, Shakhatreh, 2003, United States. Agency for International Development Usaid. Center for Population and Nutrition, 1998, Shrestha et al., 1991). These proximate determinants of fertility are not usually included in data collected by vital registration. Contraceptive prevalence is usually measured by asking survey respondents whether they are using a family planning method at the time of the survey. Increasingly, contraceptive dynamics are also measured through the use of event history calendars (Belli, 1998, Freedman et al., 1988, Callahan and Becker, 2012, Becker and Diop-Sidibe, 2003, Becker and Sosa, 1992) designed to capture method discontinuation and switching. These surveys also increasingly ask respondents about their fertility intentions, e.g., whether they want to have a child now or later. These questions permit measuring the level of unwanted pregnancies and possible unmet need for family planning.

Section III. Current estimates of proposed demographic SDG indicators in LMICS

KEY POINTS

- The available data sources in LMICs often permit obtaining direct estimates of most proposed demographic SDG indicators.
- There are however significant data gaps since these data are not available for all countries in any given year.
- Various UN inter-agencies groups and IHME use statistical models to produce estimates of proposed demographic SDG indicators for all countries.
- But these model-based estimates of SDG indicators cannot serve as "report cards" for the SDGs because:
  - They are highly sensitive to errors in model specification and underlying assumptions
  - They are affected by biases in input data
  - They are affected by "endogeneity bias", since they estimate mortality rates on the basis of information about health expenditures and income per capita. In these models, the most expensive programs would often automatically appear very successful in reducing mortality.
  - They are affected by concomitant but unrelated development processes in other countries

(a) Overview

Data availability varies significantly across countries and between proposed SDG indicators. Users of the SDGs thus face considerable challenges in producing comparable estimates. Here, we describe some of the analytical approaches currently

used to produce such estimates and we highlight some of their limitations. We focus on LMICs since the estimation of proposed demographic SDG indicators in HICs poses significantly fewer problems.

There are currently two approaches to obtaining demographic estimates in LMICs. The first approach is direct: it calculates events and exposure times from independent datasets collected for the purpose of demographic estimation. Unfortunately, the application of direct measurements is limited by the availability of demographic datasets in LMICs. The second approach, on the other hand, is model-based. It draws on statistical inferences to a) combine multiple data sources, and b) generate estimates of demographic rates for countries-years for which no data are available.

(b) Direct estimates

(i) Prospective estimates from vital statistics

In a number of LMICs, estimates of mortality and fertility rates are obtained directly from vital registration data. In the 2012 World Population Prospects, for example, estimates of fertility rates were obtained directly from vital registration for most countries in Latin America and several North African countries (e.g., Algeria, Tunisia). In sub-Saharan Africa, only the estimates for South Africa were derived from vital registration data, whereas in south Asia, this was the case only for India (sample vital registration system) and Sri Lanka. Mortality estimates for these countries were also produced using vital registration data, except for South Africa for which models were used to account for the mortality impact of the AIDS epidemic.

(ii) Prospective estimates from longitudinal studies

In LMIC settings where longitudinal demographic studies are ongoing, estimates of demographic rates are also available for the small populations undergoing surveillance. In particular, the sites of the INDEPTH network routinely produce estimates of cause-specific mortality rates, which are often difficult to obtain from vital registration data. INDEPTH thus recently released comparative studies of HIV (Streatfield et al., 2014c, Reniers et al., 2014), Malaria (Ndila et al., 2014, Streatfield et al., 2014b) and pregnancy-related mortality (Streatfield et al., 2014a) in its sub-Saharan and south Asian sites (see Figure 6). In some countries (e.g., Ghana, Kenya), HDSS have been strategically placed so that they provide data on the major ecological or cultural regions

of a country. Most HDSS are however located in rural areas, and thus do not provide information on the living conditions and demographic processes affecting urban populations. Two notable exceptions include the Nairobi and Ouagadougou HDSS (Rossier et al., 2014a, Rossier et al., 2014b, Rossier et al., 2012)

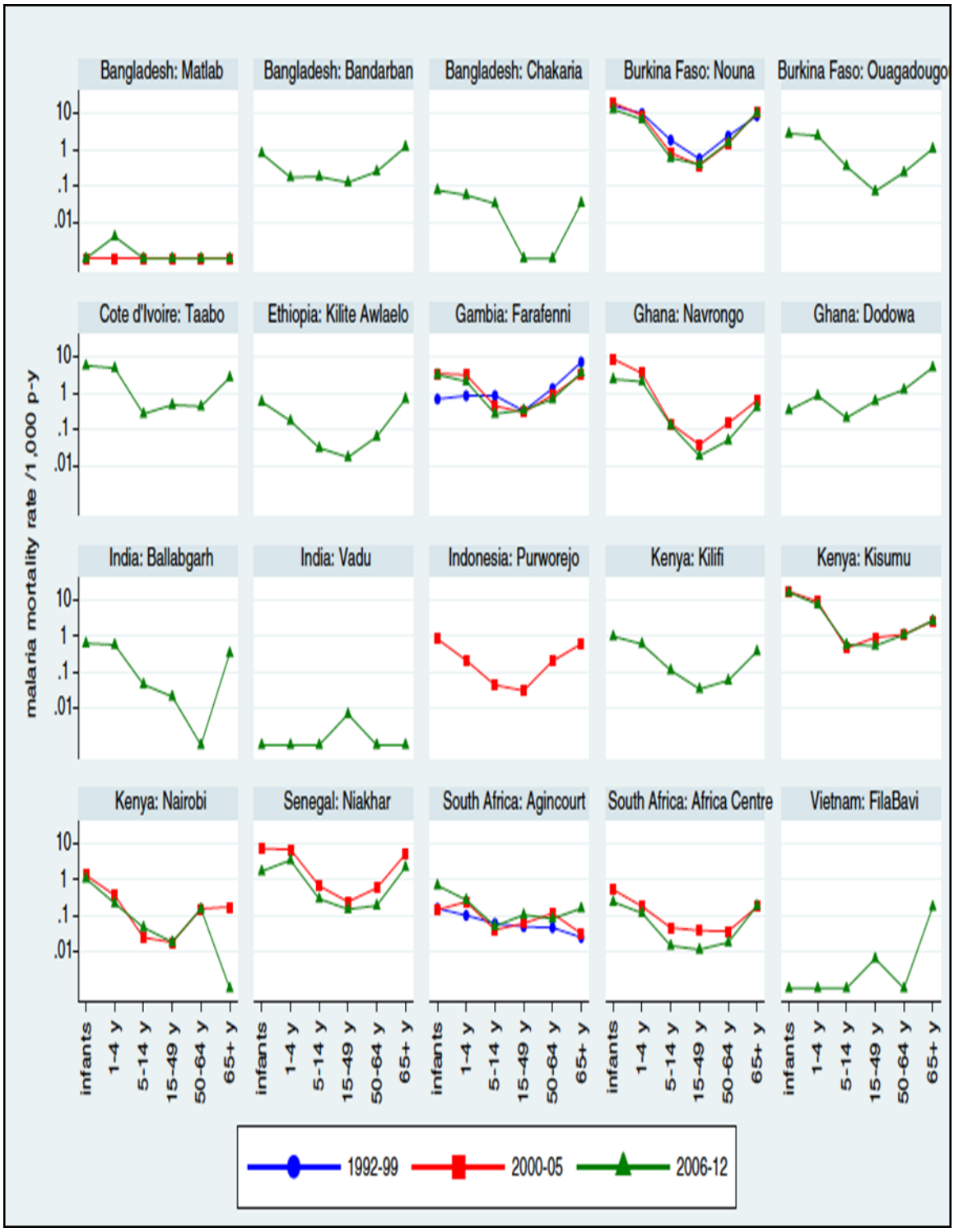

Figure 6: trends in malaria-mortality at sites of the INDEPTH Network, from Streatfield et al. 2014a

(iii) Retrospective estimates

In LMICs, estimates of the proposed mortality and fertility SDG indicators can also be directly and independently obtained from some of the alternative data sources described

above, particularly survey and census data. These estimates are retrospective, since respondents are asked to recall the births and deaths that have occurred over a recent past.

Among the proposed demographic SDG indicators, estimates of maternal mortality, neonatal mortality, infant mortality, road traffic deaths, violent deaths, fertility rates and contraceptive rates can all be obtained from a single dataset. For example, an estimate of the maternal mortality ratio (MMR) would be obtained from DHS/WHS/MICS data using SSH as follows. First, the numerator of maternal mortality rates is obtained by counting the number of pregnancy-related deaths among the sisters of a respondent who died at ages 15-49 years old. Second, the denominator of maternal mortality rates is calculated using information on the dates of birth and death of a respondent's sisters. Finally, the MMR is obtained by dividing the maternal mortality rate by estimates of the general fertility rate derived from birth histories collected during the same DHS (Moultrie et al., 2013, Preston et al., 2001). Similarly, childhood mortality rates are calculated from reports of dates of birth and death of children under 5 obtained during birth histories. DHS and MICS survey reports thus frequently include estimates of such rates, as do census reports.

In producing estimates of proposed SDG indicators, survey and census datasets have numerous limitations:

1) They cannot independently produce estimates of several indicators including death rates from HIV, Malaria or TB. This is the case because whereas survey data on maternal deaths or violent deaths have reasonable sensitivity/specificity (Helleringer et al., 2013), this not the case for HIV or Malaria-related deaths. These deaths require much more extensive VA data for precise classification. Other approaches are thus needed to produce estimates of these indicators, which frequently involve complex multi-stage modeling exercises (see below).
2) They often require very large sample sizes to produce reliable estimates of mortality rates. This is so because deaths (especially when separated by cause) are rare events. It is thus not rare for the sample size of mortality surveys to exceed 20,000 households. For surveys with smaller sample sizes, the confidence intervals attached to mortality estimates are very large and thus not informative (Hill et al., 2006, AbouZahr et al., 2007, Hill et al., 2007)

3) Linked to the previous limitation, survey datasets do not allow producing annual time-series of mortality and fertility indicators of interest: instead estimates are only available for relatively long time periods (e.g., 5-7 years for maternal mortality, 3-5 years for childhood mortality) preceding the survey. This is the case because births and deaths remain rare events, i.e., on the order of 10-30 per 1,000 person-years. As a result, adequate sample sizes can only be reached by pooling together several years of retrospective reports.
4) Survey datasets on mortality and fertility are affected by a series of errors and biases, which significantly affect resulting estimates. If the direction and magnitude vary from country to country, then survey estimates of mortality and fertility may not be comparable across population, even if similar questionnaires were used to obtain the data. These errors include:
    - Age heaping: mortality rates and life-table quantities require precise data on age, but individuals in LMICs often only have limited knowledge about their date of birth. As a result, they often report age figures that are rounded to the nearest multiple of 5 (see Figure 7). When reporting events within childhood, heaping may also happen at 7 days for neonatal deaths (Pullum et al., 2013, Pullum, 1991) or at 6 and 12 months (for infant deaths).
    - Reporting errors: such errors occur when respondents report BH and/or SSH that differ from the true survival of their children and/or siblings (Helleringer et al., 2014a, Helleringer et al., 2014b, Stanton et al., 2000). They may be due to recall issues, poor wording of questions or interviewer behaviors. They have repeatedly been found to bias survey estimates of mortality and fertility rates.
    - Sample selection biases, which occur because mortality data are typically obtained from a sample of survivors. Families with zero survivors are not included in the survey (Gakidou and King, 2006). In addition, the likelihood of inclusion in a survey may be related to the risk of dying. This may lead to bias in estimates if, for example, survival is associated with the number of siblings in a family, or if the survival of mother and children are correlated, as in the case of the HIV epidemic (Masquelier, 2013, Obermeyer et al., 2010, Hallett et al., 2010).
    - Social desirability biases, which occur when survey respondents do not disclose stigmatized events or behaviors. This concerns particularly reports

of sexual behaviors (Cleland et al., 2004, Helleringer et al., 2011, Mensch et al., 2003, Mensch et al., 2008a, Mensch et al., 2008b, Hewett et al., 2008). But it may also affect reports of fertility (e.g., respondents not reporting an out-of-wedlock birth during the birth history) or mortality (e.g., respondents not reporting the death of a sibling due to HIV/AIDS or another stigmatized cause).

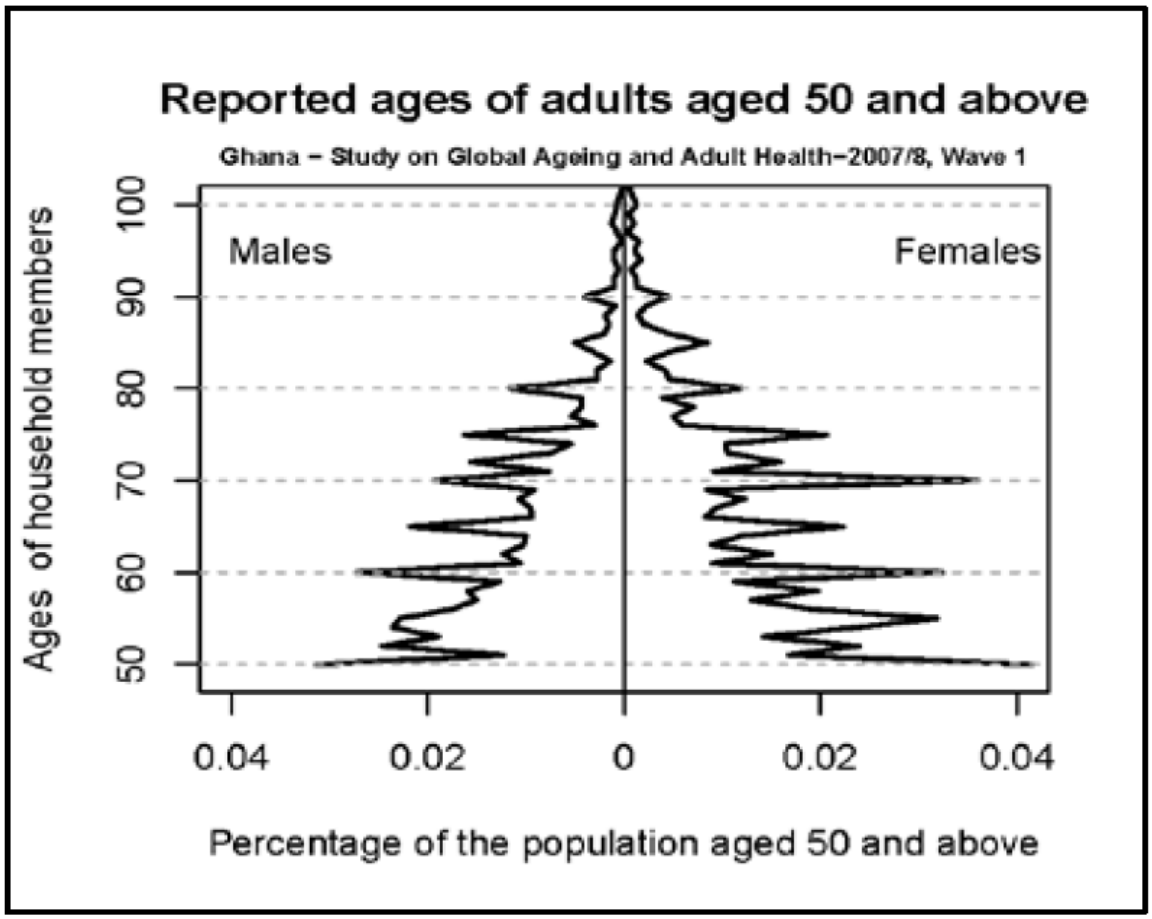

Figure 7: Example of Age Heaping among respondents aged 50+ years old during WHO SAGE survey in Ghana

(c) Filling in the gaps: model-based estimates

Direct estimates can thus currently only be obtained for a subset of country-years in LMICs due to limited data availability. Various UN inter-agency groups and the Institute

for Health Metrics and Evaluation (IHME) thus use statistical models to “fill in” the data gaps and obtain estimates for countries/years for which data are not available.

(i) General approach

These models proceed in multiple stages, as follows (Kassebaum et al., 2014, Wang et al., 2012, Lozano et al., 2011, Hogan et al., 2010, Obermeyer et al., 2010, Mathers et al., 2004, Mathers et al., 2002, Mathers et al., 2001). First, all available data sources on a particular indicator in each country are obtained and pooled. These datasets are then used to produce preliminary estimates. Second, these estimates are adjusted/corrected in multi-country regressions, which include a number of covariates such as GDP per capita, health expenditures per capita, coverage of specific health interventions etc… Finally, regressions estimates are used to obtain estimates of mortality and fertility rates for which no data are available through out-of-sample predictions.

In other words, if no data for malaria mortality are available for, say, DR Congo, then an estimate will be produced on the basis of a) a series of country-level covariates describing DR Congo, and b) the level of malaria mortality in other countries for which mortality data are available and with characteristics similar to DR Congo. This approach has provided new information about trends in mortality in LMICs and has enabled various global studies of mortality risks.

It has also allowed obtaining estimates of healthy life expectancy (HLE) in LMICs for which data on morbidities and functional limitations are not available (i.e., the global Burden of Disease Study). Estimates of HLE are obtained through the same process of data amalgamation and extrapolation. First, analysts systematically review the links between various risk factors (e.g., risky sex behaviors, smoking, alcohol use) and mortality/morbidity in longitudinal, small-scale epidemiological studies. Then, they obtain data on the prevalence of these risk factors (e.g., HIV prevalence, anemia) for mortality and morbidity in each country. Finally, they extrapolate the prevalence of functional limitations in a country on the basis of the assumed relations between risk factors and disabilities. Once estimates of the prevalence of disabilities/morbidities have been obtained through this multi-stage process, standard techniques of HLE calculation can be applied.

Only a small number of studies have collected more precise data on the prevalence of disabilities and functional limitations in selected LMICs. For example, the WHO organized the Study on Global aging and adult health (SAGE), which collected data on functional limitations and health among individuals aged 50 and over in a number of LMICs in Africa, Latin America and Asia (Kowal et al., 2012). These studies permit calculating HLE through techniques similar to those used in HICs, e.g., the Sullivan method (Harttgen et al., 2013, Payne et al., 2013).

(ii) Can model-based estimates serve as SDG report cards?

Until recently, global estimates of mortality/fertility indicators for all countries have been produced at relatively widely spaced intervals, e.g., every 5-10 years or so. To respond to the increasing demand for more frequent estimates however, IHME recently moved to annual updates of its GBD study. Updated estimates for each country can indeed be obtained each year by fitting similar models to new datasets that have recently been collected. We argue however that this model-based strategy should not constitute the primary approach to monitoring demographic SDG indicators, for several reasons.

- First, the external validity of the estimates obtained during the GBD and other global studies of mortality/fertility is not ascertained. Instead, these studies primarily aim to produce estimates that are internally consistent. This is particularly problematic since these estimates often rely on survey data as their primary data input, and these data are affected by known biases and errors (see above). More frequent updates of GBD estimates would only serve to propagate such errors and biases from one year to the next (serial correlation).
- Second, estimates for earlier years can often be revised *a posteriori*, thus possibly leading to new conclusions about progress towards mortality targets or the efficiency of development programs. This is particularly problematic in the SDG context, since indicators should act as “management tools” to help (re)orient programs on an ongoing basis.
- Third, the estimate of indicator *Y* in country *C* may change from year-to-year even if no new data have been collected in *C* on *Y*. This would happen because estimates of *Y* are obtained through multi-country multivariate regressions. As a result, *Y* may change if a) some of the covariates describing country *C* change

from one year to the next, or b) the levels of *Y* in countries similar with covariates similar to those observed in *C* change from one year to the next. These models may thus lead to spurious trends in countries where no improvement in mortality are observed.

- Finally, these models may suffer from severe endogeneity bias, i.e., mortality outcomes are estimated on the basis of data on inputs whose effectiveness we are trying to evaluate. Specifically, final estimates of mortality rates are obtained through regression models, which include health spending per capita, the coverage of various health interventions, or GDP as covariates. If country *C* increases its health spending, or invests in insecticide-treated nets (ITN) to prevent Malaria, then estimates of mortality rates in *C* will be adjusted downwards to match those of other countries with similar levels of health spending or ITN coverage. The estimates of mortality rates are thus not independent of the interventions/programs SDG indicators are supposed to evaluate. Such model-based strategies thus cannot be used as “report cards” for health and development programs.

Section IV. Ensuring that SDG indicators serve as report cards for health and development programs

KEY POINTS

- We propose to organize the proposed demographic SDG indicators hierarchically, following existing theoretical frameworks used to study fertility and mortality in demography and population studies.

- We warn that both life expectancy and the total fertility rate are complex indicators, which result from the interactions between distal and proximate determinants. They cannot serve as report cards for any particular health programs or even the health sector as a whole. They are best conceived as inter-sectorial indicators.

- We identify several issues associated with proposed SDG indicators measuring the proximate determinants of fertility/mortality, including:
  - Misalignment of the measurement and program timeframes
  - Associations between causes of death
  - Ill-defined populations at risk
  - Presence of confounders
  - Poor predictive value of data on fertility intentions & contraceptive use

- Several strategies can help address these issues, including:
  - Redefining several indicators (e.g., cause-specific death rates)
  - New and/or modified data collection protocols (e.g., incorporating measures of strength of fertility intentions, contributing causes)
  - Collecting extensive residential and migration data to account for changes in the composition of national and local populations
  - Adopting standardization techniques for monitoring trends in indicators net of changes in population composition

Sub-Section 4.01 A hierarchical SDG monitoring framework

In order to ensure accountability in the health and development sphere, the proposed SDG indicators must serve as “report cards” for the health and development programs that address them (SDSN, 2015). Development actors should however only be held accountable for the indicators their actions and interventions can independently modify. This is not the case of all indicators proposed so far. Some indicators are directly influenced by program interventions, and are well suited to serve as SDG indicators. But the dynamics of other indicators are the result of complex interactions between multitudes of independent factors. Trends in these latter indicators may be difficult to attribute to the actions of any single development actor. In addition, the negative effects of the actions of some actors may offset the beneficial effects of the actions of other development actors on such indicators. As a result, they are not well suited to act as “report cards” for any particular interventions or even for the health sector as a whole.

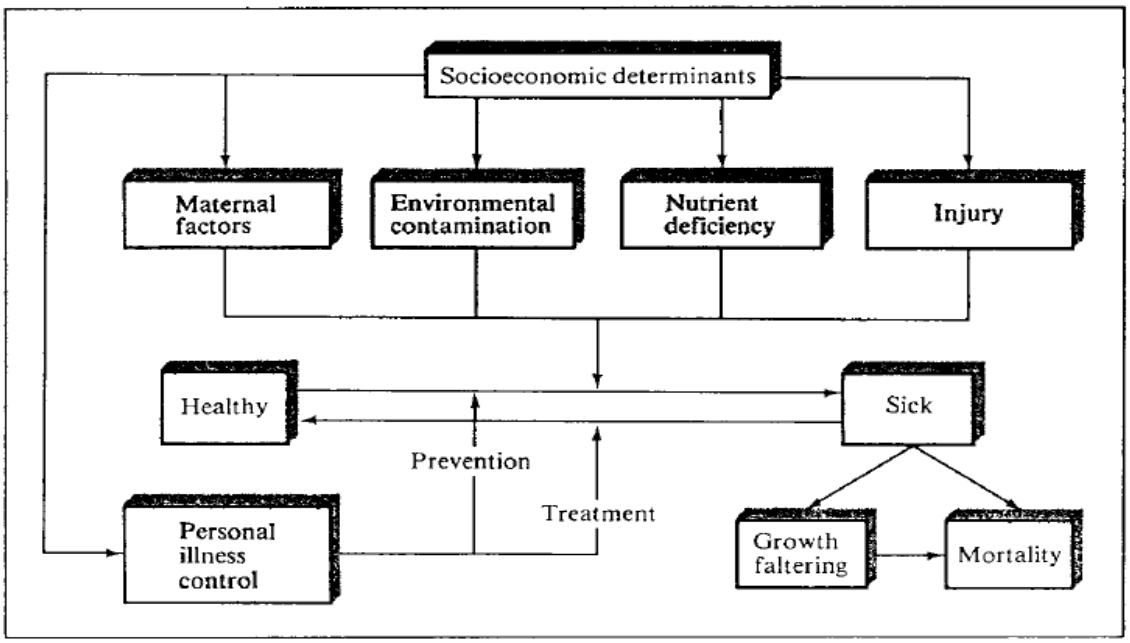

Figure 8a: Proximate determinants framework for the study of under-5 mortality developed by Mosley and Chen (1984)

In demography, the total fertility rate and life expectancy are two measures that involve such complex interactions. Both are determined by several proximate determinants, as well as a number of more distal causes. Proximate determinants are the immediate biological and behavioral factors, which influence human fertility/mortality. Distal causes

include socio-economic, cultural, or historical factors that also influence human fertility/mortality, but only through their impact on the proximate determinants. The proximate determinants of fertility, for example, include age at first marriage, contraceptive use, abortions, or breastfeeding behaviors (Bongaarts, 1978, Kalule-Sabiti, 1984). Mosley and Chen (1984) outlined the theoretical framework required to assess the proximate determinants of child mortality (see Figure 8a), and another framework of the proximate determinants of HIV infection has been outlined more recently (Boerma and Weir, 2005).

These hierarchical frameworks highlight the fact that the effects of any single proximate or distal determinant on the level of fertility/mortality depend on the level of all the other proximate determinants (Stover, 1998, Hobcraft and Little, 1984, Bongaarts, 1978). In the context of fertility, for example, the effects of an abortion on the total fertility rate depend on the level of contraceptive use. In populations where contraceptive use is high, the effect of abortions on the total fertility rate will be high, since abortions are less likely to be followed rapidly by a new conception. In populations with low contraceptive prevalence, on the other hand, the effects of abortions on the total fertility rate may be small, only increasing birth intervals by a few months on average.

In the context of SDG monitoring, this implies that both the total fertility rate and life expectancy cannot serve as report cards for individual health and development programs. Since their distal determinants also include socioeconomic determinants (e.g., poverty, schooling), they also do not adequately serve as report cards for a country's health sector taken as a whole. Instead, some of the trends in fertility and life expectancy may also be attributable to the performance of the educational sector, economic growth and/or the establishment of social protection schemes.

Recommendations: based on these insights from the proximate determinants frameworks, we make the following suggestions for the SDG monitoring framework.

- First, **we suggest that the total fertility rate and life expectancy should be considered as inter-sectorial SDG indicators**, rather than as report cards of the performance of specific programs or even of health system performance.
- Second, we suggest that only indicators measuring proximate determinants of fertility/mortality be used as "report cards" for specific health and development

programs in the SDG monitoring framework. Most of the demographic SDG indicators proposed so farconstitute indicators of such proximate determinants. This is the case, for example, of all the fertility and marriage indicators (contraceptive prevalence, marriage before age 18, unmet need for family planning), and of several of the mortality-related indicators.

- The proposed fertility-related SDG indicators however do not include all the proximate determinants of fertility. There is one major omission: the rate of induced abortions in a population. We thus propose that the abortion rate be included in the list of possible SDG indicators. This would permit a more thorough understanding of the determinants of fertility trends, as well as the effects of health and development programs on these determinants. In measuring abortion rates and monitoring whether health and development programs ensure greater access to safe reproductive health services, distinctions should be made between safe and unsafe abortions. The WHO defines unsafe abortion as a "procedure for terminating a pregnancy performed by persons lacking the necessary skills or in an environment not in conformity with minimal medical standards, or both" (Ganatra et al., 2014). Measurement procedures for unsafe abortions have been proposed (Adler et al., 2012, Gerdts et al., 2013, Ganatra et al., 2014), which often use a combination of data from health facilities and surveys. Such methods should however be strengthened to limit biases and permit more accurate measurement of trends in abortion practices.
- Similarly, the indicator on breastfeeding (Percentage of children less than 6 months who are fed breast milk alone), which helps assess the duration of post-partum amenorrhea in fertility models, could be improved upon. It could instead be reframed as the average duration of breastfeeding among recent mothers. This indicator is readily calculated from data already collected during large population-based surveys like the MICS or the DHS. It presents the advantage of being much easier to incorporate into models of the proximate determinants of fertility. It also helps control for differential censoring by breastfeeding status between children who have died before the survey and children who have remained alive.
- Further theoretical work is needed to produce a list of the proximate determinants of mortality from various causes and at different stages of the life cycle. Current frameworks focus on narrow age ranges (e.g., children under-5) or specific

diseases (e.g., HIV). Extended frameworks are needed to capture NCD mortality, as well as mortality from Malaria and TB.

The pyramidal framework illustrated in Figure 8b summarizes these recommendations. At the bottom of the pyramid, process indicators (e.g., number of activities conducted, staff employed etc…) are collected on a continuous basis. Data on the proximate determinants are then collected among the target population every year or so (see below) and serve as report cards. Finally, data on demographic outcomes (i.e., total fertility rate and life expectancy) are collected every 3-5 years and are related to data on proximate determinants analytically.

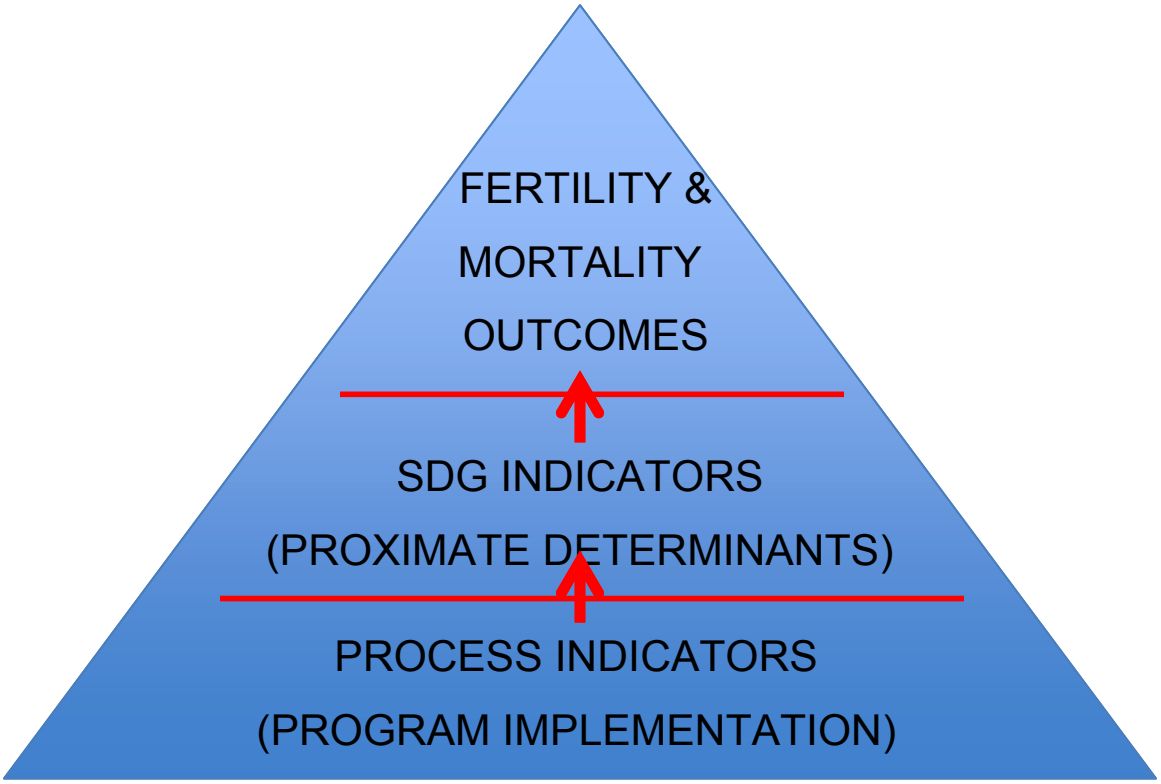

Figure 8b: Schematic illustration of the proposed hierarchical SDG monitoring framework

Sub-Section 4.02 Strengthening proposed demographic SDG indicators

(a) Criteria for SDG indicators to serve as "report cards"

We now turn to examining each of the individual demographic indicators proposed by the SDSN. We evaluate their ability to serve as report cards for specific health and development programs. In that perspective, it is important that the proposed indicators:

- Are likely to be associated with interventions conducted by health and development programs,
- Capture the total effects of health and development programs on a demographic process, rather than a subset of the causal pathways through which these programs influence wellbeing; and
- Permit measuring these effects net of the influence of possible confounders.[2]

The SDG indicators that are currently proposed do not meet those criteria. We identify several weaknesses associated with these indicators and we suggest solutions to address them.

(b) Accounting for assocations between causes of death

A crucial problem in measuring the impact of health and development programs on mortality concerns the potential for multiple contributing causes of death. Typically, analyses of mortality focus on the underlying cause recorded on a death certificate. But in a number of instances, several other causes may have contributed to the death (Desesquelles et al., 2014, Fedeli et al., 2014, Fink et al., 2012, Pacheco et al., 2011, Tardon et al., 1995, Wong et al., 1978). This is particularly so in the case of HIV: HIV infection significantly increases the incidence of other diseases (e.g., TB, pneumonia, non-communicable diseases) and may aggravate other conditions (e.g., pregnancy-related conditions). The incidence of Diabetes mellitus is also increased among HIV

[2] The primary approach to controlling for the effects of confounders in program evaluation entails conducting randomized controlled trials of specific interventions. However, we focus on the development of indicators for national programs; and at that level RCTs are rarely, if ever, the appropriate evaluation strategy. It will thus likely be impossible to account for all confounders in the SDG monitoring strategy. Our aim here is to point the potential role of major confounders, which can be ruled out through simple redefinitions of the proposed indicators and/or statistical adjustments.

patients who are taking protease inhibitors (Tien et al., 2007, Justman et al., 2003). Finally, HIV-related deaths may often be attributed to other concomitant causes on death certificates (Dorrington et al., 2002, Naghavi et al., 2010, Phillips et al., 2014).

In that context, monitoring cause-specific death rates in isolation may misrepresent mortality changes. Associations between causes of death may significantly confound estimates of the impact of health and development programs. For example, with the advent of antiretroviral therapy, HIV patients may no longer die from AIDS-related illnesses, but may instead be at an increased risk of dying at later ages from diseases whose incidence is heightened by HIV infection (e.g., various cancers, kidney disease) or by HIV treatment (e.g., Diabetes). In such settings, a significant proportion of an observed increase in the number of deaths due to NCDs may also be attributable to HIV.

Box 1: Model-based approaches to assessing associations between causes of death, the case of pregnancy-related and HIV deaths

The HIV epidemic led to a rapid increase in reproductive age mortality.

Among women with HIV who died during pregnancy or within 42 days of a delivery, which deaths were due to maternal causes? And which deaths were solely related to HIV infection?

To answer these questions, IHME proceeds in several steps.

1) It reviews all available prospective epidemiological studies on the risk factors for pregnancy-related deaths
2) From these studies, it obtains a pooled estimate of the relative risk of pregnancy-related death associated with HIV infection.
3) Then, it uses this estimate in combination with estimates of HIV prevalence generated independently by UNAIDS to assess the % of pregnancy-related deaths that are due to HIV in each country.

The UN MMEIG used a slightly different approach to this problem, possibly explaining divergence in estimates of maternal deaths between the two groups.

A similar issue affects current estimates of maternal mortality: HIV has significantly accentuated the levels of reproductive-age mortality in southern and eastern African countries, leading to a large increase in the number of deaths occurring during or shortly after pregnancy (Zaba et al., 2013, Calvert and Ronsmans, 2013, Le Coeur et al., 2005). In the absence of adequate data on multiple causes of death, complex statistical models are used to decipher what proportion of this increase is attributable to maternal causes vs. HIV alone (Box 1).

In HICs and in some LMICs, the issue of association between causes of death is often addressed by collecting data on multiple contributing causes of death on death certificates (Fink et al., 2012, D'Amico et al., 1999, Tardon et al., 1995). Then analysts use these data to detect associations between causes of death that are more frequent than expected. This often yields surprising results: in some HICs, for example, the multiple causes of death (MCOD) approach led to reevaluating the role of diseases of the blood and diseases of the skin in mortality (Desesquelles et al., 2014).

In LMICs for which VAs are the primary sources of data on causes of death, it is also possible to ascertain multiple causes of death, even though most studies have focused on attributing the underlying cause of death. Multiple causes of death can be attributed using VA data either by giving specific instructions to physician reviewers, or by statistical modeling (King and Lu, 2008). These approaches are not however widely adopted and instead single-cause files are extracted from VA data.

Recommendations: To address issues raised by associations between causes of death, we recommend that:

- When not included, VR forms should be revised to systematically include the possibility of recording multiple causes of death.
- VR forms should also systematically include a check box to record the HIV status of the decedent, if known.
- VA questionnaires should systematically inquire about the HIV status of the deceased, including whether the deceased was tested and on ARV treatment;
- VA data should be analyzed using statistical models for assigning multiple causes of death assignment, rather than simpler models currently in use.

These new data will permit measuring the associations between causes of death empirically. They will permit including all deaths caused (directly or through association) by a disease on the report cards of health and development programs that target it. This will considerably improve over current model-based estimates of the interaction between important causes of death (see box 1).

(c) The timeframe of program effects

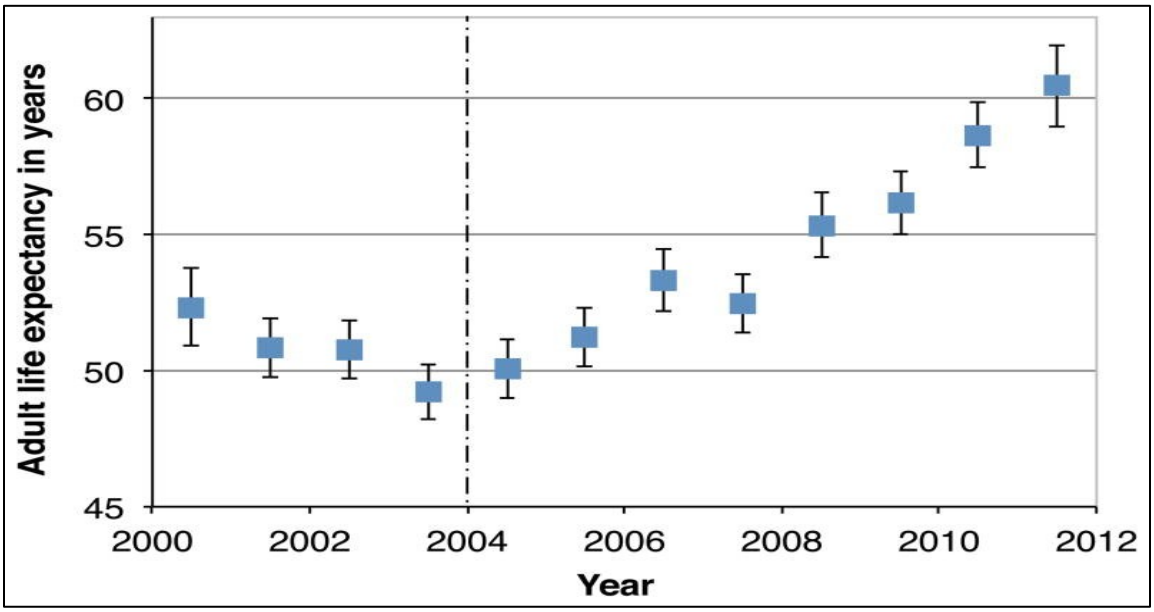

Figure 9: effects of the rollout of antiretroviral therapy in KwaZulu-Natal, South Africa. From Bor et al. 2013

For an indicator to serve as "report card", we also need to formulate a theory about the amount of time it will take for the potential effects of programs to unfold and modify an indicator. In some programs, these effects may be almost immediate, e.g., ARV programs (see Figure 9) or anti-malarial distribution programs (Bor et al., 2013, Trape et al., 2012). In other programs, on the other hand, the expected effects may take significantly longer to unfold. This is the case of programs addressing non-communicable diseases: whereas some components of NCD programs may have immediate impacts on mortality (e.g., strengthening NCD care in hospitals, surgery programs), other components may have delayed effects that unfold over time. Smoking prevention programs, or interventions targeting obesity, for example, may not translate into significant mortality reductions immediately.

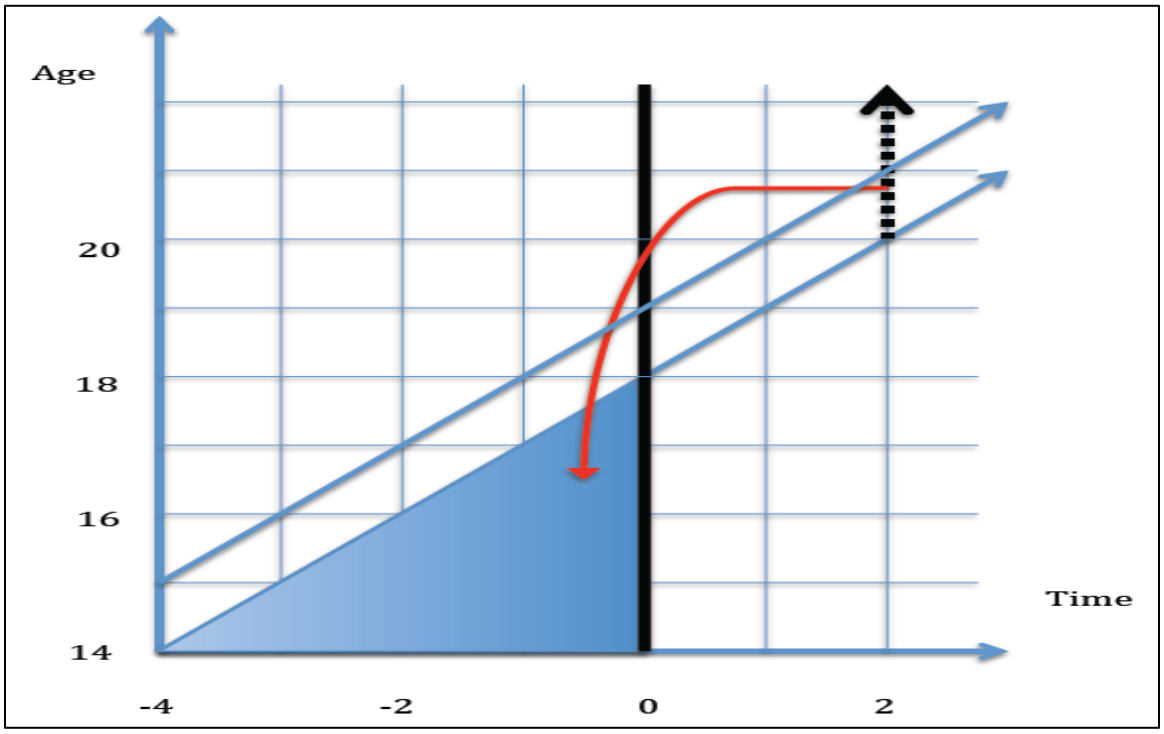

Figure 10: illustration of the discrepancy between the timeframe of the proposed marriage indicator and the timeframe of possible program effects

Notes: The vertical black line represents the start time of program implementation, whereas the diagonal lines represent cohorts of individuals. We hypothesize that a survey is conducted at time t = 2; the sample is represented by the dotted vertical arrow. Given the definition of the indicator, this survey only includes respondents aged 20+ at that time. Respondents are asked to report whether they were married at age = 18 years old, i.e., between 2 and 6 years before the survey. As a result of this long recall period, respondents aged 20 years old at time t = 2 contribute only person-years of observation that were not affected by program implementation (represented by the blue triangle).

The proposed marriage indicator provides another example of an indicator whose measurement timeframe is misaligned with the timeframe of program effects. This indicator is defined as the proportion of 20-24 years old that were not married at the time of their 18th birthday. In the Lexis diagram in Figure 10, we represent this discrepancy: if measured at time t=2, the proposed indicator will only capture behaviors that have taken place before the program to prevent early marriage was implemented.

Recommendations: To address issues related to the timeframe of dimensions of program effects, we recommend that:

- The indicator currently proposed to measure the prevalence of child marriage (i.e., the proportion of 20-24 years old who were not married by age 18) should be abandoned. Instead, trends in early marriage should be monitored through estimation of the quartiles of the distribution of age at 1st marriage. This indicator constitutes one of the proximate determinants of fertility. It is already included in most DHS and MICS surveys through the collection of detailed marital histories. It can be also be computed by simply asking respondents their current marital status at the time of a survey. Life table techniques are then used to calculate the quartiles of the distribution of age at 1st marriage in cohorts. To monitor progress towards the elimination of child marriage (i.e., before age 18), health and development programs should monitor the number and proportion of marriages that involve a child. Such figures can be estimated from marital histories collected during the DHS and the MICS surveys.
- Changes (or lack thereof) in indicator of NCD-related mortality should only be interpreted in parallel with changes in other proposed SDG indicators describing NCD-related risk behaviors. These include, for example, indicators related to tobacco use, alcohol consumption and medical care in hospitals. Interpretation of trends in NCD-related mortality indicators should be based on a theory of change linking modifications in risk behaviors and healthcare-seeking practices to health and morbidity outcomes.

(d) Strengthening Indicators based on intentions and self-perceptions

Several of the proposed demographic SDG indicators require population members to self-report their own fertility intentions (e.g., met family planning). Such data are typically collected during surveys, e.g., DHS or MICS. Relying on reports of a respondent's perceptions may however also introduce biases in assessments of health and development programs. This may be the case if intentions are not truthfully stated during a survey due to social desirability biases. It may also be the case if the intentions are not clear and can vary over time and across different types of social interactions (Pearson and Becker, 2014, Speizer et al., 2013, Hayford et al., 2012, Withers et al., 2011, Hayford and Morgan, 2008)

We consider the case of fertility intentions and the ability of programs to meet the demand for family planning. The associated SDG indicator is defined as the proportion of women who do not want (more) children or who want to postpone childbearing, who are currently using contraceptive methods. The ability of this indicator to serve as a report card for family planning programs rests on the underlying dynamics of fertility intentions among women. If a woman's fertility intentions are "set in stone", (e.g., if they adopt "stopping" behaviors), then the link with health-seeking behaviors is likely direct: she will seek family planning services to implement her strong preferences. If she cannot access such services at that time, then she experiences unmet need. On the other hand, if her intentions are more fluid or are influenced by other persons (e.g., spouse, family members), then the link between her stated intentions and her health-seeking behaviors might be more tenuous (Trinitapoli and Yeatman, 2011, Bongaarts and Casterline, 2013, Kodzi et al., 2010a, Kodzi et al., 2010b, Casterline and El-Zeini, 2007, Casterline et al., 1997). Similarly, a woman may be postponing births for reasons other than child spacing/stopping (Moultrie et al., 2008, Moultrie et al., 2012). She may thus state her intentions to delay or stop childbearing during a survey interview, but this may not prompt her to seek services. In such instances, her lack of contraceptive use cannot necessarily be counted as "unmet need" and may not be attributed to weaknesses of the family planning services.

Recommendations: To remedy this limitation, surveys on fertility and family planning behaviors should incorporate probabilistic assessments of the strength of fertility

preferences. Specifically, instead of asking respondents to state whether they want to have more children or not, respondents are asked to consider how likely it is that they may have a child in the next n years. Several methods to quantify such subjective probabilities have been tested, including in sub-Saharan countries (Delavande and Kohler, 2012, Delavande and Manski, 2012, Delavande and Kohler, 2009, Trinitapoli and Yeatman, 2011). Analysts should first test whether the level of unmet need for family planning varies by the strength of underlying fertility intentions. If so, this information should then be included as weights in the construction of the unmet need indicators.

(e) Measuring the effectiveness of contraception: typical vs. perfect use

To serve as a report card, an indicator must also capture the total effects of a program on a demographic process, not simply some of the causal pathways through which a program may produce effects. In the case of the measurement of contraceptive prevalence, this criterion is not met. Currently, the indicator is defined as the percentage of women of reproductive age who use a given contraceptive method at a given point in time. This definition 1) does not consider the "method mix" among users, and 2) does not take "adherence" into account, i.e., whether women use these contraceptives consistently.

Both dimensions modify the association between the level of contraceptive prevalence and the level of fertility in a population (Stover, 1998, Chimere-Dan, 1990, Hobcraft and Little, 1984, Bongaarts, 1978). Over time, the contraceptive mix may move towards more effective methods, e.g., injectables and sterilization. In other settings, however, it may become less effective, e.g., if people move away from more permanent methods and switch to pills or condoms. Similarly, certain methods are only effective if women take them consistently, e.g., pills. In populations where such methods are highly prevalent, it will thus be particularly important to measure adherence to contraceptives in order to evaluate the effectiveness of a family planning program.

Recommendations: we recommend that the definition of contraceptive prevalence be modified to incorporate the two related dimensions of "contraceptive mix" and adherence. This will permit measuring the effective use of contraceptives, rather than

solely contraceptive uptake. Surveys like the DHS already ask questions about the type of contraceptive method that respondents use. These allow describing the method mix. Adherence, on the other hand, can be ascertained on the basis of standard medication adherence scales (for pills and condoms) or through the use of a retrospective event calendar for other methods (e.g., injectables). Such a calendar is also already included in the standard DHS questionnaire on contraceptive use. These parameters can serve to calculate a new indicator of "effective contraceptive prevalence".

Specifically, we call *W* the total number of women of reproductive age. Among those are *C* contraceptive users, so that contraceptive prevalence is simply $\frac{C}{W}$. Among the *C* contraceptive users however, there are m_i users of method *i.* There are *n* methods available in this population, so that $\sum_{i=1}^{n} m_i = C$. We call e_i the effectiveness of each method *i*, with $e_i = 1 - RR_i$ where RR_i is the relative reduction in the monthly risk of conception associated with the use of method *i,* so that $0 < e_i < 1$. RR_i can be calculated using life-table techniques and data from a contraceptive calendar or can be borrowed from previous studies (Bongaarts, 2014, Stover, 1998, Bongaarts, 1978). Then we define the new weighted indicator of effective contraceptive prevalence as: $P = \frac{\sum_{i=1}^{n} e_i \times m_i}{W}$.

This indicator should be supplemented by an ascertainment of the availability of different contraceptive methods at family planning providers. Indeed, in some instances, a highly effective contraceptive mix may actually represent situations where have few contraceptive options to choose from and their reproductive rights are constrained. This may occur for example in programs where providers focus on delivering IUDs or on promoting permanent methods of contraception like female sterilization. This ascertainment can be conducted using data from health facilities, e.g., data collected through the PMA 2020 surveys (see box below) or the Service provision assessment (SPA) surveys routinely conducted as part of the DHS.

(f) Controlling for confounders in SDG indicators

In evaluating national-level programs, most indicators will also entail potential for confounding. Trends in SDG indicators may thus not capture changes in the risk of

occurrence of demographic events but may instead simply reflect changes in the composition of the population at-risk.

To illustrate this point, we consider the example of deaths from road traffic injuries. This indicator is included to help ascertain the effects of road traffic safety programs in preventing premature deaths. As such, an increase in the number of traffic-related deaths per 100,000 inhabitants should be perceived as a sign of poor performance from these programs, which should then be held accountable. However this is not the case: the interpretation of that indicator is complicated by concomitant trends in car ownership and utilization of public transportation. Over the next 15 years in most LMICs, we should expect the number of road traffic deaths per 100,000 people to increase significantly simply because more and more individuals will gain access to cars and other transportation modes. It is also expected that an important number of roads will be built in LMICs. In such context of expansion of the road network and traffic, an increase in the number of road traffic deaths does per 100,000 inhabitants does not necessarily imply that roads are becoming increasingly dangerous and that road safety programs are failing. It may instead only reflect the fact that more and more individuals are exposed to risk of traffic injuries/deaths.

Similar issue arises in the case of maternal mortality, and other cause-specific mortality indicators (e.g., mortality associated with HIV, TB and Malaria). Maternal deaths are particularly concentrated at older ages and higher parities (Nove et al., 2014). During the course of a fertility transition, such births will become increasingly rare however because women are often adopting “stopping” behaviors, i.e., decide not to have any more children after they have already had two or three. This change in the composition of births may prompt declines in the maternal mortality ratio, but this decline is not related to improvements in programs providing access to obstetric care. Instead it may be attributable to family planning programs. Similarly, increases in the death rate associated with HIV, TB or Malaria may not necessarily be due to lower quality of patient care, but instead may also be related to i) increases in the size of population groups most vulnerable to these diseases, and ii) changes in care-seeking behaviors among affected populations.

Finally, migration can also be a pervasive confounder in monitoring SDG indicators and using these indicators as report cards. Both in and out-migrants often present very different characteristics relative to the baseline population of interest. For example, there are a number of studies highlighting migration-related "paradoxes" in health, i.e., the migrant population is healthier than the baseline population even though it may often have lower socio-economic status (Palloni and Morenoff, 2001, Franzini et al., 2001, Palloni and Arias, 2004, Patel et al., 2004, Smith and Bradshaw, 2006, Turra and Goldman, 2007, Ho et al., 2007, Drummond, 2011, Pinheiro et al., 2011, Borrell and Lancet, 2012, Thomson et al., 2013, Bostean, 2013, Young and Hopkins, 2014b, Young and Hopkins, 2014a). Such "healthy migrant" bias may confound the evaluation of health programs. This was shown recently in an evaluation of the role of municipal health programs in increasing life expectancy in New York City: in that context, most of the differential trends in life expectancy between NYC and the rest of the US were due to the relative health of immigrant populations in NYC (Preston and Elo, 2014). Other forms of selective migration can also confound SDG indicators. For example, out-migrants are often more educated than the rest of the population. As a result, some SDG indicators that are correlated with educational level may decline in sending communities during times of intense out-migration. Both international and internal migration flows can confound the monitoring of SDG indicators and the evaluation of the effectiveness of health and development programs.

Recommendations: To account for potential confounders in the proposed demographic SDG indicators, we propose several modifications/extensions to the definition of these cause-specific indicators.

1. <u>The indicator related to traffic deaths should be redefined as the number of traffic deaths per vehicle-kilometer.</u> In a number of countries, data on the number of kilometers travelled per year per vehicle are often available from road safety authorities, or can be obtained through observation at short, regular intervals (Abegaz et al., 2014, Bhatti et al., 2011, Sobngwi-Tambekou et al., 2010, Lagarde, 2007). It constitutes the standard way of measuring risk associated with road traffic in epidemiology.
2. <u>For indicators focused on Malaria, HIV and TB, it may be helpful to include a secondary indicator defined as the number of deaths per year among cases of the diseases.</u> This case fatality ratio would permit controlling for trends in

incidence and would help focus on the effectiveness of health services in a) diagnosing and b) treating these illnesses.

We do not recommend changes to the definition of the maternal mortality indicator. Instead,

3. We strongly recommend the use of standardization techniques to control for confounding factors. These techniques are commonly used in demography and should be routinely incorporated in the SDG monitoring toolkit (Preston et al., 2001). They permit “removing” the effects of changes in population composition on crude rates. These techniques are also helpful because they can possibly be extended into decomposition models, which will be helpful for disaggregation purposes.

Finally, in order to account for the potential confounding arising from migration flows,

4. We recommend systematically collecting information on the residential history of survey respondents over the past several years before a survey. This information will help control for changes in the composition of populations targeted by health and development programs. Standardizations and decomposition techniques similar to those described in point #3 immediately above can help remove the effects of migration from trends in SDG indicators and capture the effects of health and development programs on these indicators.

Section V. Can we produce high-quality annual time-series of the proposed demographic SDG indicators?

KEY POINTS

- For SDG indicators to serve as management tools/report cards, the frequency of data collection on demographic processes must increase in most LMICs

- High-frequency low-cost surveys of the proximate determinants of fertility and mortality, aided by the use of new technologies, provide valuable information for health programs

- But initiatives to increase frequency of mortality/fertility measurements are affected by
  - Limited validity (e.g., scale-up of births/deaths registration through community key informants)
  - Limited statistical power, since fertility/mortality remain rare events (e.g., on the order of 10-30 per 1,000 person-year)

- For the foreseeable future, the measurement of fertility/mortality will thus still rely on retrospective survey data in most LMICs

- We formulate the following recommendations:
  - Resources permitting, countries should adopt a system of dispersed demographic surveillance, similar to the SAVVY project in Tanzania
  - Other LMICs should adopt a tiered data collection system, with yearly surveys of proximate determinants of fertility/mortality, followed by large fertility/mortality studies every 3-5 years
  - Methodological research on strategies to improve quality of retrospective survey data on fertility/mortality should be intensified.
  - New innovative strategies to monitor migration (e.g., geocoded data, big data from cellphone communication, social networking sites) should be investigated.

In addition to the conceptual modifications detailed above, monitoring the proposed demographic SDG indicators will require increasing the frequency of data collection in most LMICs. The SDSN calls for annual reporting on the various SDG indicators, except healthy life expectancy, total fertility rate and maternal mortality rate. We have already ruled out earlier the possibility of annual reporting using estimates obtained from statistical models. This approach would likely generate considerable bias in monitoring of SDG progress. Instead, the monitoring of SDG indicators should be based on empirical data. In this section, we identify which demographic SDG indicators can be reported annually, and we articulate a possible data collection framework for LMICs. This framework matches the (theoretical) proximate determinants framework for mortality/fertility analysis described in the preceding section. The first tier of that framework consists of annual data collection (e.g., surveys) on the proximate determinants of mortality and fertility. The second tier consists of large-scale surveys every 3-5 years to obtain estimates of a) mortality by cause and b)

(a) Initiatives to increase the frequency of data collection

Annual reporting of SDG indicators constitutes a significant departure from the MDG era, when a number of indicators were only reported every few years or possibly never at all. This ambitious goal is however facilitated by the rapid introduction of new technologies in population-based surveys. We review the effects of these technologies. We then describe how these technologies permitted expanding existing data collection initiatives (e.g., HDSS, DHS) or developing new data collection approaches.

(i) The role of new technologies

In recent years, demographic data collection in LMICs has increasingly moved from paper-based data collection to electronic supports and paperless tools. In particular, a fast-growing number of surveys use mobile phones or tablets for data collection, building on software platforms like the Open Data Kit (ODK). These new technologies permit strengthening demographic data collection in several important ways:

- New technologies significantly reduce survey costs: the use of these new tools permits efficiency gains at several steps of the survey process. First, moving from paper-based to electronic data collection induces savings related to printing, rental of storage space and physical archival of paper questionnaires. Second, they bypass the need for large data entry teams, which are often required to

conduct double data entry for quality assurance. These costs are all variable costs, which need to be incurred at every round of data collection. By comparison, the cost of purchasing and programming mobile phones/tables for data collection can be considered a fixed cost: it is incurred once during the first round of a survey, then additional costs only include maintenance and improvements.

- New technologies permit closer monitoring of data quality: once data are collected, they can be transmitted through the GPRS network to the central server of the data collection organization. There, these data are checked and cleaned using pre-defined routines. Feedback is then provided to study interviewers and supervisors, who can seek further information, re-contact study respondents and/or address data quality issues “on the spot”.
- New technologies significantly reduce the amount of time between data collection and availability of study results. Once all data collection has been completed, the dataset is immediately available for final cleaning and data analysis. Survey reports can thus be produced much more rapidly, i.e., in a few weeks vs. several months or years for paper-based surveys.
- New technologies permit increasing data confidentiality and privacy. Whereas paper-based surveys are vulnerable to questionnaire loss and theft, data collected electronically can be protected by passwords, encrypted, and transmitted through secure websites. Electronic data collection also reduces the number of individuals who handle the data during a survey, thus reducing the likelihood that someone will inadvertently disclose information. Finally, some new technologies (e.g., audio computer-assisted self-interview, ACASI) permit bypassing interviewers altogether. In ACASI, the respondent hears pre-recorded survey questions directly through headphones, and then answers these questions on his/her own by keying in answers on a keypad. ACASI has often reduced social desirability biases in surveys conducted in LMICs. It thus promises more accurate data on sensitive subjects such as contraception, sexual behaviors or HIV/AIDS.
- Finally, new technologies also often allow rapidly collecting a number of new important data points, which were rarely available in previous paper-based surveys. This is the case in particular of geo-referenced data. Most phones/ tablets used for data collection purposes now also incorporate a GPS tracker,

which automatically records the location where data were collected. It can also rapidly obtain coordinates of a number of points of interest in local communities, e.g., health services, water points, schools. The use of new technologies thus also promises richer datasets on the determinants of fertility/mortality.

Box 2: Monitoring family planning programs through high-frequency mobile-phone based surveys

In 2012, the *Family Planning 2020* initiative was launched, which pledges to increase access to family planning for 120 million of girls and women in the poorest countries.

To enable a precise evaluation of this initiative, a program of high-frequency surveys was launched, called Performance, Monitoring and Accountability 2020 (PMA2020).

PMA2020 uses the Open Data Kit software on mobile phones to conduct rapid surveys of family planning indicators in nationally representative samples.

Between 150-200 communities are selected at random, then enumerators are posted in these communities for data collection.

Data are transmitted immediately over the GPRS network, and are checked, cleaned and collated in real time.

Then, enumerators are provided with direct feedback from the PMA2020 platform, which permits rapid dissemination of study results to participating communities

So far, PMA2020 surveys have been conducted in Ghana, Ethiopia, Nigeria, Burkina-Faso, DR Congo, Kenya and Uganda

More information on the PMA2020 survey program can be found at:
http://www.pma2020.org

(ii) Examples of high-frequency data collection

Several data collection initiatives in LMICs are currently using these new technologies to accelerate/expand data collection on demographic indicators. These include first a

number of survey initiatives. For example, the launch of ***continuous demographic and health surveys (cDHS)*** in Peru and more recently in Senegal represents a significant departure from current practice in DHS (Corsi et al., 2012). cDHS are conducted every year over a 5-year period. Each year, a team of study interviewers visits 1/5 of the sample, selected at random. This allows obtaining annual estimates of a number of indicators traditionally estimated by the DHS. By pooling the data across all 5 years, the less common events measured by the DHS can also be precisely quantified (e.g., fertility). cDHS systematically use mobile phones for data collection, which permits obtaining a full survey report only a few weeks after the completion of data collection. Several other initiatives to conduct high-frequency surveys in LMICs have so far focused on family planning (see box 2).

They also include a number of ambitious initiatives, which aim to provide an assessment of fertility and mortality rates in real-time. For example, there are several ongoing attempts to ***scale-up activities of health and demographic surveillance (HDSS)***, so that HDSS data can be representative at the national level. These projects also emulate some aspects of the design of the million deaths study conducted within the sample vital registration system of India (Million Death Study et al., 2010, Jha et al., 2006). A small number of HDSS clusters are selected at random in a country (e.g., n ≈ 30), then an HDSS is set up in each cluster, providing data on fertility/mortality rates, as well as patterns of causes of death through VA. In Tanzania (see box 3), the Ifakara health institute leads such a projects (Kabadi et al., 2014). Other projects have attempted to ***develop local systems of births/deaths monitoring***, by mobilizing either community health workers or community key informants. Specifically, following an initial census, these agents are tasked with recording vital events that occur in a community on a day-to-day basis and transmitting information about these events through new information technologies (Amouzou et al., 2014, Bowden et al., 2012, Caleo et al., 2012, Roberts et al., 2011, Roberts et al., 2010).

(iii) Big data

Several reviews have also highlighted the potential for "big data" to complement existing sources of demographic information. Big data refer to traces of digital processes and social media exchanges, which can be then be used to learn about social interactions or other "real-world" processes. Records of Internet searches, use of mobile phones, or history of purchases and transactions, for example produce big data. In HICs, big data is now frequently used in epidemiological surveillance (Wiwanitkit, 2014, Araz et al., 2014, Santillana et al., 2014, Lazer et al., 2014, Thompson et al., 2014, Dugas et al., 2013, Pervaiz et al., 2012), and several researchers have investigated the use of Big data in

Box 3: Tanzania's national platform for health impact evaluation

Tanzania is one of the countries with the most health and demographic surveillance systems (N = 5).

Recently, Tanzania also launched an ambitious project of dispersed health and demographic surveillance, aimed at providing high-quality and nationally representative data on fertility and mortality.

This sentinel panel of districts (SPD) selected 23 districts at random, and in each district, installed a HDSS designed to provide continuous monitoring of fertility and mortality.

This SPD also provides a unique opportunity to conduct a number of nested health evaluation studies, including randomized evaluations of new health policies or interventions.

The SPD complements, rather than replaces, other information systems already in place at the district or national levels, including survey programs and routine health information systems.

The SPD is also planned as a learning platform for the countrywide scale-up of vital registration activities.

fertility research. In LMICs, big data has also been explored for the monitoring of outbreaks of infectious diseases, particularly the spread of Malaria (Wesolowski et al., 2014, Buckee et al., 2013, Wesolowski et al., 2012) or the recent Ebola outbreak in West Africa (Mandl, 2014, Milinovich et al., 2014). There have been few applications of big data in fertility and mortality research in LMICs, so far, however. This is primarily the case because such data rarely also include information on key covariates in fertility/mortality research, e.g., age and sex. Instead, big data sources often only provide signals about aggregate trends only (e.g., number of events, cases of a disease). Migration may be a more promising area for the application of big data in demographic measurement in LMICs. For example, data from mobile phone operators can be used to track movements within countries (Wesolowski et al., 2013, Wesolowski et al., 2014), whereas data from social networking sites like LinkedIn have been used to monitor international migration (State et al., 2014).

(b) Issues associated with high-frequency data collection

Despite the enthusiasm generated by the use of new technologies and by big data in demographic data collection, several limitations persist when we try to increase the frequency of demographic data collection. These limitations need to be taken into account when formulating SDG monitoring plans. We detail these issues below.

(i) Statistical power and variance of estimates

The first issue is related to statistical power and concerns particularly the measurement of demographic events like births and deaths. The SDSN already ruled out yearly reporting on life expectancy, maternal mortality, total fertility and NCD mortality, because of a) limited year-to-year change, and b) huge sample size requirements for precise measurement of annual change (SDSN, 2015). We argue that this also applies to age-specific (e.g., neonatal mortality) and cause-specific (e.g., malaria mortality) death rates. Indeed, in most LMICs, crude death rates and crude birth rates are comprised between 10 and 40 per 1,000 person-years. In other words, in order to “find” 100 deaths during a data collection exercise, one must collect information on up to 10,000 persons. Childhood deaths or malaria deaths constitute subsets of all deaths, so that finding 100 such deaths requires listing even more people.

	Sample size			Events in past 12 months		
	Households	Household members	Women aged 15-49	Births	Infant deaths	Under-5 deaths
2009	26,988	105,225	24,212	2,047	38	248
2010	26,605	101,409	22,947	1,706	27	197
2011	26,528	98,662	22,517	1,664	26	196
2012	27,218	103,211	23,888	1,807	25	175

Table 2: Sample sizes available during the 2009-2012 rounds of the Peru Continuous DHS

We show in table 2 the number of demographic events listed during the continuous DHS in Peru, i.e., a nationally representative survey conducted each year, between 2009 and 2012. During this survey, a high number of households (>25,000) were visited each year, yielding data on approximately 100,000 household members. Despite this high sample size, each survey recorded data on only 25-40 infant deaths and 175-250 under-5 deaths per year. Such figures yield insufficient statistical power to detect year-to-year variation in childhood mortality even at the national level. Statistical power is even more limited to investigate population dynamics within sub-groups, as is recommended for SDG monitoring. In countries with higher mortality rates among children (e.g., Senegal, another country with a continuous DHS), the statistical power to measure short-term changes in childhood mortality may be slightly higher but still limited.

Continuous surveys also create “built-in” sample size limitations. We illustrate this point in Figure 11, which depicts a typical situation where cross-sectional surveys are conducted every year during the course of a program. The goal of that hypothetical program is to prevent childhood deaths. During each survey, women are asked about their births of the past 5 years, as in the DHS. In that fictitious population, there are 100 births per year, which are captured during each survey. Then, at the end of the cycle of continuous surveys, the number of births for which data are available will be highest at the beginning of the program, but will be lowest for the final year of the program. This data collection design will a) require the development of sampling weights to account for the under-representation of births in years 3-5, b) make the assessment of program

impact very uncertain and c) may introduce bias if program effects occurred with a lag (i.e., were largest for years 3-5).

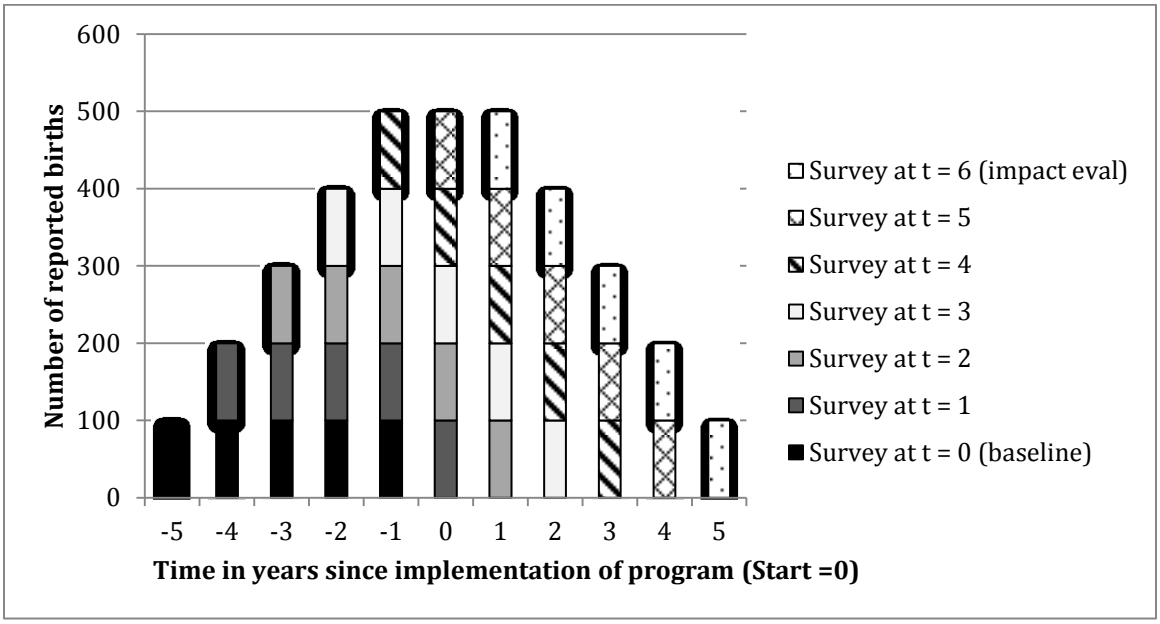

Figure 11: Hypothetical sample sizes available after multiple cross-sectional surveys for the evaluation of a program to reduce child mortality.

Notes: a survey is conducted each year in January. In this population, there are 100 births per year and this has not changed over time. We assume that all births over the past 5 years are reported during the survey, i.e., there is no recall or reporting errors.

Whereas such surveys yield inadequate power to measure changes in demographic rates, they often yield very adequate power to detect short-term changes in the determinants of these rates. This is the case, for example, of measures of contraceptive prevalence or of measures of health-seeking behaviors among mothers of children under 5.

(ii) Explanatory power and Bias

The second issue concerns the validity of the estimates produced by high-frequency surveys and other modes of annual data collection. Some of the data produced by these initiatives may be imprecise, particularly those that rely on recall of past events or the reporting of current status data. They are collected using the same questionnaires used during the DHS and other surveys. They are thus affected by the same biases we listed

above including reporting errors, sample selection biases and social desirability biases. Increasing the frequency of such surveys does not address the issues that limit the underlying quality of demographic data.

Even the methods based on prospective monitoring of demographic events suffer from a number of potential biases. For example, the development of local systems of birth/death registration has often resulted in fairly large under-estimates of childhood mortality rates. This is the case because community health workers and/or key informants miss some of the deaths that occur in their community (Amouzou et al., 2014, Amouzou et al., 2013). HDSS may also fail to record a number of demographic events and/or misclassify causes of death (Pison et al., 2014, Jha, 2014, Delaunay et al., 2013, Ye et al., 2012, Oduro et al., 2012).

Some datasets resulting from high-frequency surveys have so far only been partially validated. For example, data from repeated family planning surveys (i.e., PMA 2020) have only been compared with data from the DHS to ensure that figures were globally consistent with previous estimates of contraceptive prevalence. They have not been validated against, for example, facility-based medical records of contraceptive uptake, nor have they been tested in models of the proximate determinants of fertility. Similarly, data from dispersed HDSS (e.g., SAVVY project in Box 3) have not yet been fully validated. It is thus unclear whether they maintain the same high quality as data collected in smaller HDSS despite more limited supervision and less frequent update rounds.

The persistence of these biases in high-frequency demographic datasets presents significant challenges for the monitoring of SDGs. On the one hand, it is not clear whether these biases are time-invariant or are instead affected by time-varying factors, which may be related to health and development programs. For example, following an intensification of family planning programs, survey respondents may be more inclined to report using contraceptives or may be less inclined to report abortions during surveys. Such biases would confound the estimates of program impact on SDG progress, making programs look more effective in promoting contraception than they truly are.

On the other hand, measurement biases may also jeopardize cross-country comparisons of SDG progress. This is the case because reporting errors may vary across populations. For example, validation studies of SSH data on maternal mortality have indicated that SSH under-estimated the proportion of maternal deaths in Matlab in Bangladesh, but over-estimated this proportion in Bandafassi in Senegal (Shahidullah, 1995a, Shahidullah, 1995b, Helleringer et al., 2013)

(iii) Frequency vs. detail of data collection

Finally, tradeoffs may emerge between the frequency of data collection and the level of detail contained in each dataset. Indeed, in order to enable more frequent survey rounds, questionnaires may often be stripped down to include solely a number of key questions. These questions are those that are needed to produce estimates of the key SDG indicators. However, we mentioned earlier that, in order to serve as a report card for health and development programs, SDG indicators should also be measured net of possible confounders. This will often require the measurement of a number of covariates, which do not directly enter directly in the calculation of SDG indicators. These covariates nonetheless play a key role in evaluation of health programs. They are incorporated as controls in multivariate models to account for trends in confounding factors. They may also be tested as possible effect modifiers. Without including measurements of these covariates, SDG monitoring may ultimately fail to attribute observed changes in SDG indicator to the effects of health and development programs. It may also fail to detect possible sub-groups for which the programs are beneficial/detrimental (see below).

(c) A proposed data collection system for monitoring of demographic SDG indicators in LMICs

Ideally, significant resources would be invested in vital registration systems in LMICs early in the post-2015 era, so that SDG progress can be measured directly. Few countries are however in a position to commit such resources to vital registration. Furthermore, because of training requirements and infrastructure development, such investments would only result in increased coverage and accuracy of vital registration data after several years. In the interim period, alternative strategies for the consistent measurement of SDG indicators are thus needed. These strategies are not substitutes

for the development of vital registration systems. They are designed to provide the information required by health and development programs, during the scale-up of vital registration coverage levels.

In some LMIC settings where resources can be mobilized, it may be useful to implement a dispersed health and demographic surveillance system, akin to the SAVVY project conducted in Tanzania (see box 3). It entails selecting at random a number of communities around the country, then conducting health and demographic surveillance in these communities. It has multiple advantages because it combines the strengths of HDSS (i.e., high-quality prospective data on vital events, continuous assessment of causes of death) with some of the strengths of survey datasets (i.e., representativity). It also permits building capacity for vital registration in a country. The drawbacks of the dispersed HDSS approach include a) possible Hawthorne effect (i.e., individuals under demographic surveillance may modify their demographic or health-seeking behaviors), b) difficulties in accounting for migration at the local level, c) costs and d) a lack of validation of key empirical estimates.

In most LMICs, the monitoring of demographic SDG indicators will primarily rely on the collection of survey datasets. Specifically, in order to track SDG progress and measure the impact of programs, we recommend that this monitoring system should three key components:

- A high-quality census conducted every 10 years,
- Annual (or even more frequent) surveys on the proximate determinants of fertility and mortality and, the prevalence of morbidities and functional limitations;
- More extensive surveys of mortality and fertility rates every 3-5 years (see Figure 12).

This proposed system is based on hierarchical theoretical frameworks of the determinants of fertility and mortality (Khan and Shirmeen, 2007, Stover, 1998, Hobcraft and Little, 1984, Bongaarts, 1978), which we described in the previous section. These frameworks identify a number of distal and proximate determinants, which interact to determine levels of fertility/mortality in a population (outcomes). Distal determinants include socio-demographic characteristics of households or the educational level of individuals. Proximate determinants include biological characteristics and behaviors of

individuals that directly impact their fertility/mortality (e.g., contraception for fertility, vaccinations for mortality). Both the distal and proximate determinants can be frequently measured through high-frequency surveys of reasonable size (e.g., PMA 2020 surveys on the proximate determinants of fertility). These surveys should also include measurements of self-reported morbidity and functional limitations, similar to those obtained by the WHO SAGE surveys (Leddin et al., 2013). These measures will be incorporated in subsequent measurements of healthy life expectancy.

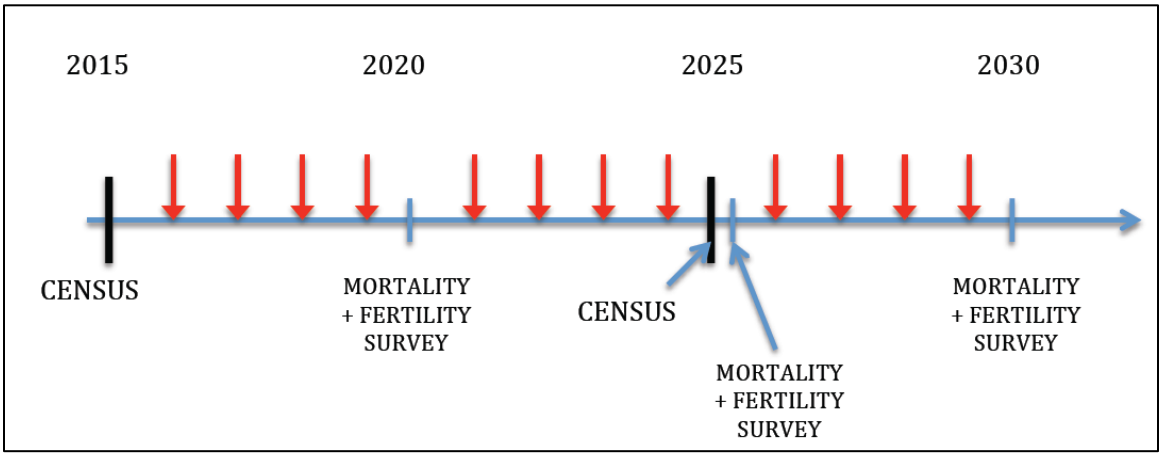

Figure 12: Data collection timeframe for the monitoring of demographic SDG indicators

Notes: red arrows indicate proximate determinants surveys to be conducted every year in target LMICs

The fertility and mortality outcomes, on the other hand, cannot currently be measured annually through surveys (see above). Instead, the proposed monitoring system captures fertility/mortality trends through large-scale surveys every 3-5 years. These surveys should use birth histories to measure child mortality, as well as siblings survival histories to measure adult mortality. They will provide snapshots of mortality and fertility levels in a population, which can then be compared over time and across countries to monitor the impact of programs.

This system ensures that annual data will be available on key determinants of demographic trends. These data will fulfill the “management tool” function of SDG indicators. The proposed data collection framework will also allow robust measurement

of mortality and fertility trends every 3-5 years, thus guaranteeing that SDG indicators can serve as report cards for health and development programs.

(d) Methodological research to support SDG monitoring

The proposed system builds on well-known methodologies. It requires however further methodological developments to ensure the consistent measurement of SDG indicators over time and their comparability across countries. We highlight three crucial areas of research.

(i) Improving quality of retrospective demographic data

The first area of methodological research focuses on improving the quality of retrospective survey data on demographic events. As we mentioned above, such data are affected by (possibly large) reporting errors, which may confound estimates of trends and program effects. Currently, most of the methodological work in this area consists of developing correction factors analytically, which can be incorporated into the complex multi-stage models used by IHME and WHO inter-agency groups to estimate mortality (Obermeyer et al., 2010).

Recently however, there have been several attempts to modify the instruments used to collected retrospective demographic data, in order to improve data quality (Helleringer et al., 2014b). This work has focused on two types of errors, which may bias estimates of mortality rates: a) omissions of specific relatives from lists established by survey respondents, and b) errors in the reporting of ages and dates at which relatives experienced various demographic events. Specifically, the inclusion of supplementary interviewing techniques (e.g., probes, recall cues and non-specific prompting) could help reduce the number of omissions in siblings' survival histories, whereas the use of an event history calendar to collect data on dates and ages at death could help improve the quality of retrospective reports. In one trial, omissions of deaths among the sisters of a respondent were reduced from 25% (using the DHS questionnaire) to 9% (using our new approach). The extent of age/date heaping, an oft-used indicator of data quality, was also greatly reduced in our new interviewing strategy (Figure 13).

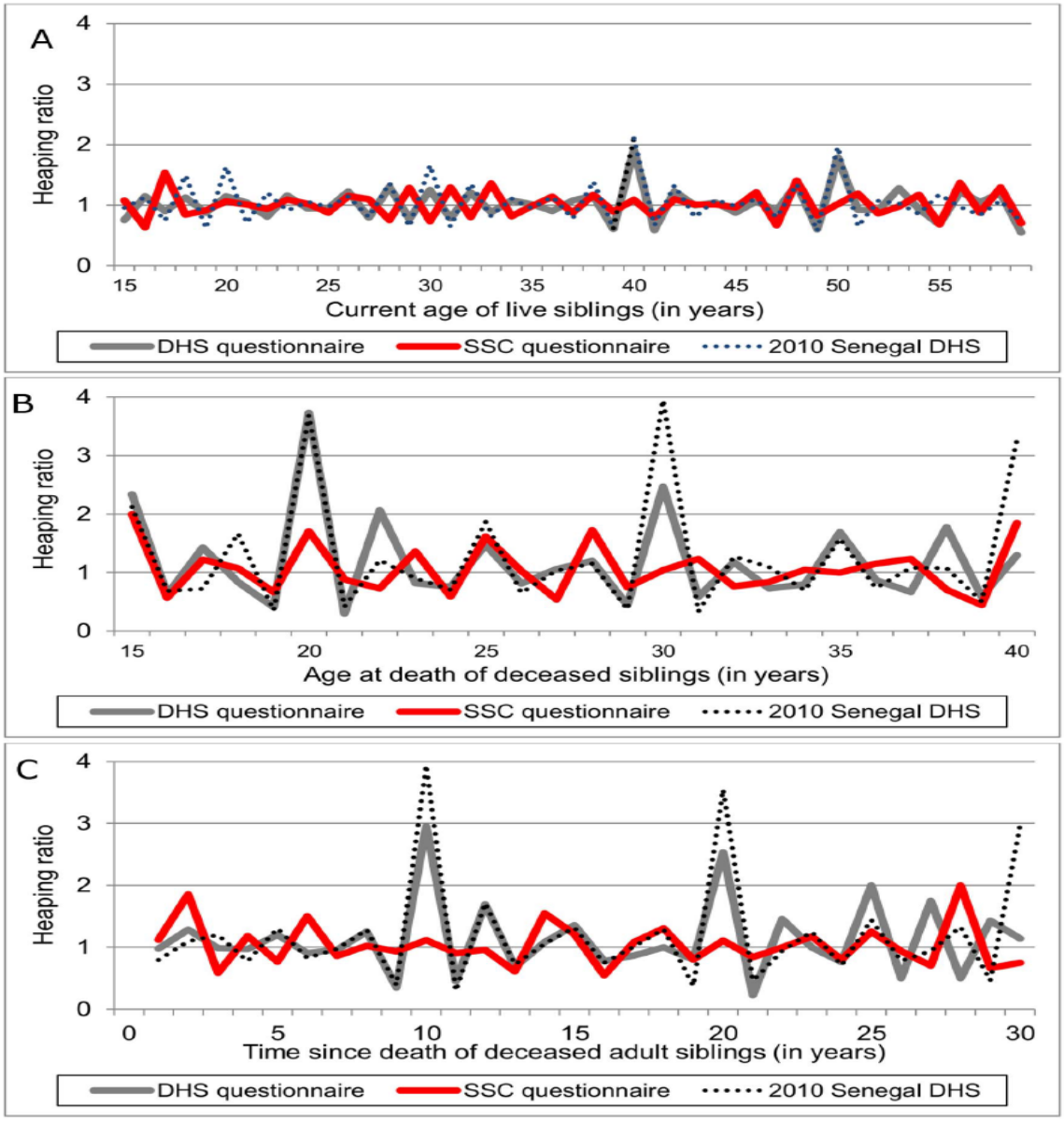

Figure 13: age and date heaping in siblings' survival histories, by type of questionnaire.

Notes: the SSC questionnaire is a new questionnaire, which includes supplementary interview techniques. These data were collected during a validation study conducted in Niakhar (Senegal) and are drawn from Helleringer et al. 2014b.

This approach to collecting retrospective mortality data should now be validated in additional settings. If it consistently improves the accuracy of data on adult mortality, it should be integrated into national large-scale surveys conducted periodically in LMICs to

measure mortality trends. Data quality improvements should also be pursued in other areas of demographic measurement:

- Trials of supplementary interviewing techniques and event calendars to improve the reporting of birth histories, including stillbirths, neonatal and infant mortality
- Use of event history calendars and other ranking techniques to accurately measure the ages of the elderlies.
- Use of supplementary interviewing techniques to help improve the reporting of deaths by cause, including limiting omissions of deaths due to stigmatized causes (e.g., HIV/AIDS, mental-health related deaths).

Several of the indicators to be collected require data on older age groups, which have seldom been the focus of demographic data collection in LMICs, particularly in sub-Saharan Africa. As during the SAGE surveys, respondents aged 50 and above must be asked about their self-perceived health limitations in order to enable measurements of healthy life expectancy. SSH, for example, should be asked to respondents asked 60 and above, so as to ascertain survival at older ages. The data provided by these respondents must however be validated in linkage studies.

(ii) Incorporating verbal autopsies in retrospective surveys

The second area of methodological development entails integrating VAs into large-scale nationally representative retrospective surveys. Whereas pregnancy-related deaths can be identified directly from SSH, without collecting VA data, this is not the case for other deaths. Deaths from HIV, Malaria and TB, in particular, are significantly more difficult to classify and require more extensive reporting of symptoms, medical history and other details of the circumstances of the death. Measuring cause-specific deaths rates during mortality surveys conducted every 3-5 years would thus entail obtaining VA information about the deaths of a) a respondent's own children, and b) his/her deceased siblings. VAs have not been systematically integrated into DHS and other surveys, so a number of additional pilots, as well as validation studies, would be required. Pilot studies would help devise standardized VA protocols for retrospective surveys, e.g., whether VAs should be collected at the end of the interview, or whether they should be integrated into the main interview; whether a smaller subset of VA question can help obtain classifications of deaths that are as accurate as those obtained from longer versions of

the VA questionnaire; and whether interviewer should seek to involve other family members rather than only asking questions to the main survey respondents. Validating VA data collected during retrospective surveys would entail for example linking DHS data with small samples of hospital deaths or larger samples of deaths obtained from HDSS.

Integrating VA data collection into large retrospective mortality surveys will require assigning the causes of large numbers of deaths on the basis of VA questionnaires. Whereas this has long been accomplished through physician review of VA data, this approach is ineffective and diverts physicians from providing care or other medical tasks. It may also introduce biases if physicians routinely over-diagnose certain conditions (e.g., Malaria). Instead, causes of death are now frequently assigned through statistical models (Ndila et al., 2014, Mossong et al., 2014, Byass et al., 2013, Misganaw et al., 2012, Ramroth et al., 2012, Vergnano et al., 2011, Fottrell et al., 2011, Tensou et al., 2010, Murray et al., 2007). These models are "trained" using a small sample of VAs for which the cause of death is known (for example from a post-mortem examination); then the patterns observed within that training sample are used as Bayesian priors in multivariate assignment models. There are multiple competing models for the assignment of causes of death from VA data, each with different strengths and weaknesses. Further statistical research is needed to identify a model, which 1) has high predictive value across most epidemiological and societal settings where VA data are collected, and 2) permits identifying associated causes of death in a multiple causes of death framework.

(iii) Improving estimates of population sizes

The final area of methodological development concerns improving estimates of population sizes, particularly in intercensal years. This is particularly important since it affects not only the demographic SDG indicators, but also all other population-base indicators, which are measured "per capita". Such indicators are often highly biased because estimates of population are highly inaccurate. In health programs, it is for example highly common to observe estimate of vaccine coverage well above 100%. This type of errors occurs because the denominator of coverage rates is obtained from

census data projected forwarded using strong assumptions about fertility, mortality and migration. Ideally, this issue should be addressed through the development of strong vital registration systems. In the interim period during which such systems are put in place however, a multi-pronged strategy is needed. On the one hand, new and improved analytical techniques are necessary to obtain reasonable projections of small-area populations (to be used as denominators in a number of SDG indicators). These techniques could build on Bayesian techniques recently adapted for population projections by researchers at the University of Washington (Alkema et al., 2011, Raftery et al., 2012, Raftery et al., 2013, Raftery et al., 2014b, Raftery et al., 2014a) and adopted by the UN population division (Gerland et al., 2014). These techniques permit quantifying the uncertainty associated with population projections. They could be extended to apply to small areas and/or population sub-groups, and should be validated using data from consecutive censuses.

On the other hand, more attention should be paid to measuring migration within countries and across borders between two censuses. As we have discussed above, migration flows may confound the monitoring of SDG indicators because it changes the composition of a population in possibly selective ways. This may threaten the ability of SDG indicators to act as report cards. But migration may also affect the accuracy of population size estimates. It is thus necessary to devise new approaches to measure migration, which may rely both on administrative data and big data derived from cellphone communications or the use of social networking sites. Such data may help complement refined analytical techniques to project small-area populations during intercensal years.

Section VI. Monitoring demographic SDG indicators in population sub-groups

KEY POINTS

- The commitment to measuring SDG progress in population sub-groups ("no one left behind") is one of the key improvements of the SDG monitoring framework over MDG monitoring

- This disaggregation is however complicated by:

  - Large sample size requirements for the detection of differences in trends between population sub-groups
  - Potential differences in data quality and accuracy across sub-groups, which may confound estimates of differentials
  - The lack of a simple statistical framework to determine whether health and development programs are reducing vs. increasing inequalities across sub-groups in a population

- To enable the disaggregation of SDG trends between population sub-groups, we recommend:
  - Surveys should be powered to detect differences in trends between sub-groups, rather than differences in the prevalence of one indicator at one point in time.
  - Methodological research should focus on measuring and addressing differences in data quality between population sub-groups
  - Simple statistical tests based on a decomposition framework should be adopted to determine convergence in demographic indicators between sub-groups.

(a) Disaggregation framework

Whereas MDGs were focused on trends in indicators at the national level, the SDGs also include a focus on reducing inequalities within and across countries. In order to ensure that “no one is left behind”, SDG monitoring frameworks must thus entail plans for “data disaggregation”, i.e., measuring trends not only at the country level but also within population sub-groups. This will permit determining whether specific population sub-groups are benefiting disproportionately from health and development programs. It will also permit refining interventions and messages, so that they reach all sub-groups and produce their expected impact.

The goal of our proposed disaggregation framework is to differentiate empirically between multiple dynamic scenarios of convergence (declining inequality) vs. divergence (growing inequality) between population sub-groups. Consider a population in which there are two groups, one advantaged with respect to a specific SDG indicator and one (relatively) disadvantaged with respect to that indicator. Then, over time, the nine scenarios summarized in table 3 may emerge. Scenarios A, B and C correspond to situations in which inequalities between population sub-groups relative to specific demographic indicators are declining (“convergence”). However, in situations B and C, the trend in this demographic indicator may be deteriorating at the country-level, depending on the size of population sub-groups. Similarly, scenarios G, H and I correspond to situations in which inequalities between sub-groups are increasing (“divergence”). Finally, scenarios D, E and F are situations in which both sub-groups co-move: inequalities are neither increasing nor declining, and national-level trends are driven by factors that are common to both population sub-groups.

		Trend among members of the advantaged group		
		Improvement	**No change**	**Deterioration**
Difference in trend between advantaged and disadvantaged groups	**Negative (i.e., convergence)**	A	B	C
	No difference (i.e., co-movement)	D	E	F
	Positive (i.e., divergence)	G	H	I

Table 3: convergence/divergence scenarios emerging during the course of health and development programs

This framework constitutes a first step in measuring trends in inequalities within a population. From a programmatic standpoint, it permits quickly identifying situations in which inequalities may be growing (e.g., scenarios G, H and I) so that they can be rapidly addressed. Subsequently, the patterns of inequality within population may be characterized more precisely using various indicators such as the Gini coefficient or the dissimilarity index.

(b) Ensuring adequate statistical power for disaggregation

This disaggregation framework first requires survey samples large enough to detect differences in trends between population sub-groups. This is an important change, relative to current practice in planning population-based surveys such as the DHS. Indeed, those surveys are typically powered to measure various indicators (e.g., maternal mortality ratios) at the country-level and to detect differences in other indicators (e.g., under-5 mortality rates) between sub-groups *at one point in time* (or during a specific period). They are not however explicitly planned to detect differences in trends over time between population sub-groups. This is problematic because often these "difference-in-differences" are much smaller (and hence more difficult to detect) than

cumulative differences measured at one point in time. As a result, two consecutive DHS surveys may, for example, not be large enough to detect, for example, a 10% difference in the pace of mortality decline between the richest and the poorest households.

In the SDG monitoring framework proposed here, on the other hand, it will be important to ensure that sample size calculations focus on detecting differences in trends between population sub-groups. This planning approach should apply both to the interim surveys of proximate fertility/mortality determinants, as well as to the larger fertility/mortality surveys conducted every 3-5 years. It will permit identifying the various scenarios depicted in table 3.

In conducting such sample size calculations, several parameters will thus need to be taken into account. These include a) the number of population sub-groups, b) the variation in demographic SDG indicator within and across sub-groups, c) the expected differences in trends between sub-groups and d) the timeframe over which this difference in trends is expected to occur. The number of sub-groups should be determined through consultation with key stakeholders, as recommended by the SDSN. (SDSN, 2015) At the minimum, SDG indicators should however be disaggregated by gender and by age. Preliminary measures of the variation in SDG indicators within and between sub-groups can be obtained from prior DHS or MICS surveys. The expected differences in trends between sub-groups should be set during consultations with key stakeholders. These differences may constitute targets the programs aim to achieve, or lower bounds in terms of inequality reduction deemed acceptable by stakeholders. The timeframe over which these differences in trends are expected to materialize should also be assessed during these consultations. It should be realistic, i.e., take into account proposed program budgets as well as the program's ability to reach the most disadvantaged groups. Survey planning for the proposed SDG monitoring framework should thus involve a number of key partners, rather than only demographers and statisticians.

(c) Accounting for differences in data quality across groups

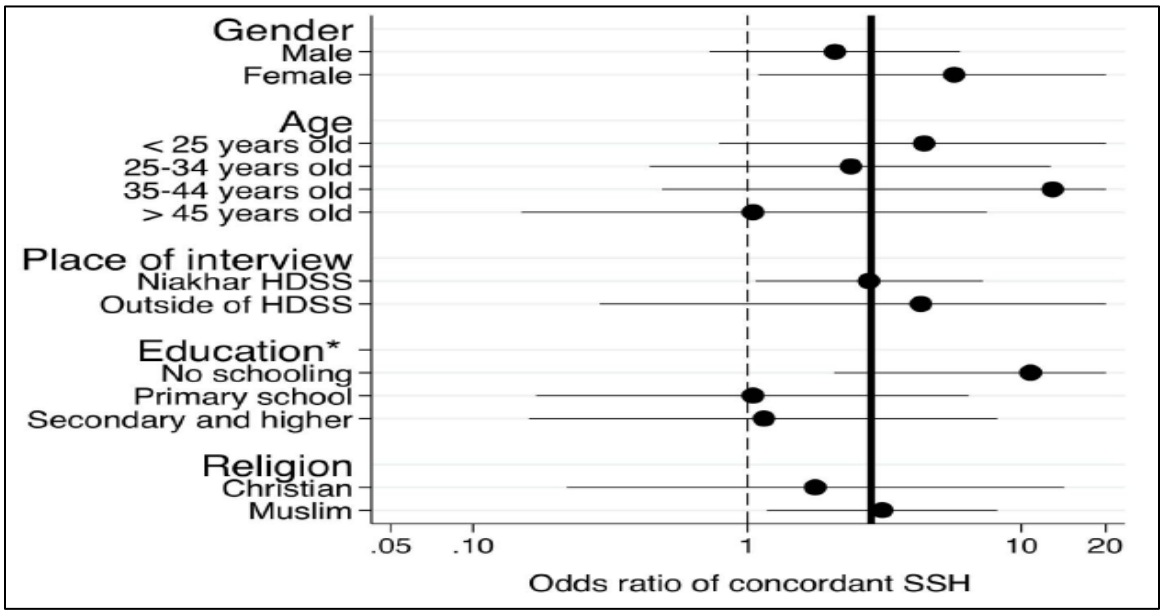

Figure 14: likelihood of reporting a concordant (i.e., correct) SSH associated with the use of the SSC questionnaire in Niakhar (Senegal).

Notes: the SSC questionnaire is a new questionnaire, which includes supplementary interview techniques, as well as an event history calendar. These data were collected during a validation study conducted in Niakhar (Senegal) and are drawn from Helleringer et al. 2014b. The dotted line represents an OR of 1. ORs above 1 indicate that the SSC collects better data than the DHS for those groups, whereas an OR below indicates the opposite. The thick black line represents the sample average.

The disaggregation framework, second, should identify and try to limit or control for differences in data quality across population sub-groups. Such differences can indeed confound results of the disaggregation: if the most disadvantaged group also reports births or deaths less completely than the advantaged group during a survey, then estimates of inequalities in fertility/mortality rates will be lower than expected. Such differential reporting errors may be particularly common when comparing groups defined by educational level or socioeconomic status.

It will thus be important to conduct tests of data quality separately by sub-groups prior to disaggregating indicators. These tests may include, for example, measures of age and date heaping, indicators of the extent of missing data, consistency in reported family size, or the frequency of impossible events sequences in birth histories. Results from such diagnostics should systematically accompany presentations of SDG indicators disaggregated across population sub-groups.

It will also be recommended to test new data collection procedures that reduce differences in data quality across the population sub-groups of interest. For example, during a trial of a new SSH questionnaire in Niakhar (Senegal), the use of supplementary interviewing and event calendars disproportionately improved data quality among the least educated respondents (Figure 14). This may be because such respondents experience the most difficulties in answering closed-ended questions, which typically form the DHS questionnaire. Other approaches to reducing gaps in data quality across population sub-groups should also be investigated, including ACASI and other new interviewing techniques.

(d) Measuring convergence/divergence in SDG indicators

Finally, we propose a statistical framework to measure convergence/divergence in SDG indicators and identify contexts defined by table 3. This framework builds on demographic decomposition techniques, first developed by Das Gupta and later refined by Chevan and Anderson (Chevan and Sutherland, 2009, Gupta and Census, 1993, Gupta, 1991, Preston et al., 2001).

We consider the case of a population with two groups similar to the population described in table 3. We have data from two surveys for these two groups. We are interested in disaggregating trends in indicator *P,* which measures for example the contraceptive prevalence. The measure of *P* in the advantaged group at time 1 is noted $P_{1,A}$ whereas the measure at time 2 is noted $P_{2,A}$. The size of the advantaged group at time 1 is noted $S_{1,A}$ whereas it is $S_{2,A}$ at time 2. The subscript D indexes the disadvantaged group.

The decompositions we propose first try to answer the following counterfactual question: how much lower/higher would *P* (i.e., contraceptive prevalence) have been in this

population if only the contraceptive prevalence among the (dis)advantaged group had been allowed to change? To answer this question, we use techniques of standardization: we calculate a counterfactual contraceptive prevalence rate, noted $P_{2,A}$, by using the population composition at time 1, i.e., $S_{1,A}$ and $S_{1,D}$ and maintaining contraceptive prevalence at its level of time 1 among the disadvantaged group. On the other hand, contraceptive prevalence among the advantaged group is allowed to change to its observed level, $P_{2,A}$. Then $P_{2,A} - P_1$ is the contribution of the advantaged group to the overall trend in contraceptive prevalence. If $P_{2,A} - P_1 > 0$, then the situation of the advantaged group is improving. In a second step, we calculate $P_{2,D} - P_1$ in a similar manner. If $P_{2,D} - P_1 > 0$, then the situation of the disadvantaged group is also improving over time.

Our test of convergence vs. divergence in contraceptive prevalence between the two groups consists of examining the difference-in-differences, i.e., $P_{2,A} - P_1 - (P_{2,D} - P_1)$. If this difference is positive, this quantity indicates that the gap in contraceptive prevalence between the advantaged and disadvantaged groups is increasing. On the other hand, if this difference is negative, it indicates that the gap is narrowing and the contraceptive use of both groups is converging. Taken together, all these measures permit rapidly assessing whether a program is promoting overall improvements as well as inequality reduction, or whether that program is solely benefiting specific sub-groups of the population. The basic functioning of these tests can be extended to be multiple groups as well as to indicators measured as continuous variables.

To test whether $P_{2,A} - P_1 - (P_{2,D} - P_1)$ is significantly different from 0, we propose to use a bootstrap approach. Specifically, the survey data set used to calculate $P_{2,A} - P_1 - (P_{2,D} - P_1)$ will be resampled *n* times, by dropping a randomly selected cluster each time. Within each bootstrap, we will recalculate $P_{2,A} - P_1 - (P_{2,D} - P_1)$. We will then examine the distribution of $P_{2,A} - P_1 - P_{2,D} - P_1$ across all bootstrap samples. Inference tests will consist of ascertaining the 2.5th and 97.5th percentiles of this distribution to determine whether $P_{2,A} - P_1 - P_{2,D} - P_1$ is significantly different from 0 (two-tailed tests).

Section VII. CONCLUSIONS

Contrary to the MDG era, ambitious plans to monitor the SDGs are being made prior to adoption of the set of goals and targets, and the launch of SDG programs. This greater attention to evaluation activities is an important step in ensuring that data will be available to guide interventions and ensure the accountability of development actors. In this report, we reviewed preliminary plans to monitor a series of demographic SDG indicators.

We found, first, that a number of modifications and re-conceptualizations were needed to help accomplish the goals of the SDG indicators, i.e., to serve as management tools and report cards for development programs. We found that a conceptual distinction was required between measures of fertility/mortality outcomes (i.e., life expectancy and total fertility rate) and indicators of the proximate determinants of these outcomes. Only the latter can serve as report cards for specific development and/or health programs, whereas measures of outcomes are best thought of as inter-sectorial performance indicators. This distinction is grounded in the proximate determinants frameworks that demographers use to study fertility and mortality, and which inform our recommendations for data collection and analysis. We also suggested that a number of proposed SDG indicators should be modified because their definition does not enable them to serve as report cards for specific development and/or health programs. These modifications entailed re-defining the population at risk (e.g., road traffic deaths), specifying the timeframe of measurement (e.g., early marriage), or weighting specific behaviors by their efficacy in achieving certain outcomes (e.g., contraception).

Second, we argued that calls to increase the frequency of data collection and reporting on SDG indicators should be qualified. Whereas data could be collected every year on the proximate determinants of both fertility and mortality, it is not currently possible to collect data annually on the actual levels of fertility and mortality. This is the case because fertility and mortality remain rare events (e.g., 10 to 40 per 1,000 person-years), which exhibit limited year-to-year variation. Detecting annual change in those outcomes

would thus require huge sample sizes and would be extremely costly. On the other hand, the proximate determinants of fertility and mortality are much more prevalent and lend themselves to repeated yearly surveys. Such surveys are further facilitated by the adoption of new technologies in survey research.

Third, we emphasized the need to conduct methodological research on survey data collection approach. Indeed, most data on mortality and fertility in LMICs is likely to originate from surveys, since vital registration systems remain incomplete. It is thus crucial to identify strategies to limit reporting errors, address selection biases and increase the confidentiality of survey data collected in LMICs.

Fourth, we noted that accomplishing the ambition of the SDG monitoring framework will require collecting new and improved data on migration flows within and across countries, and incorporating these data into existing analytical frameworks. The use of these new migration data will help 1) control for possible confounders from selective migration flows, and 2) permit obtaining precise estimates of population size (i.e., the denominator of a number of population-related SDG indicators).

Finally, we outlined a rigorous framework for measuring inequalities between population sub-groups in the various SDG indicators. In particular, we emphasized the need to simplify equity analyses by supplementing existing continuous measures of inequality (e.g., Gini coefficient) with more qualitative tools based on tests of the difference in indicators between groups. We outlined a bootstrap approach for conducting such tests when only survey data are available.

The strategies and tools described in this report represent interim strategies, which are designed to enable SDG monitoring in the absence of vital registration data. They should be considered temporary, and should be accompanied by concomitant investments in vital registration systems of LMICs.

Section VIII. REFERENCES

ABEGAZ, T., BERHANE, Y., WORKU, A., ASSRAT, A. & ASSEFA, A. 2014. Road traffic deaths and injuries are under-reported in Ethiopia: a capture-recapture method. *PLoS One,* 9**,** e103001.

ABOUZAHR, C., CLELAND, J., COULLARE, F., MACFARLANE, S. B., NOTZON, F. C., SETEL, P., SZRETER, S., MONITORING OF VITAL EVENTS WRITING, G., ANDERSON, R. N., BAWAH, A. A., BETRAN, A. P., BINKA, F., BUNDHAMCHAROEN, K., CASTRO, R., EVANS, T., FIGUEROA, X., GEORGE, C. K., GOLLOGLY, L., GONZALEZ, R., GRZEBIEN, D. R., HILL, K., HUANG, Z., HULL, T. H., INOUE, M., JAKOB, R., JHA, P., JIANG, Y., LAURENTI, R., LI, X., LIEVESLEY, D., LOPEZ, A. D., FAT, D. M., MERIALDI, M., MIKKELSEN, L., NIEN, J. K., RAO, C., RAO, K., SANKOH, O., SHIBUYA, K., SOLEMAN, N., STOUT, S., TANGCHAROENSATHIEN, V., VAN DER MAAS, P. J., WU, F., YANG, G. & ZHANG, S. 2007. The way forward. *Lancet,* 370**,** 1791-9.

ADLER, A. J., FILIPPI, V., THOMAS, S. L. & RONSMANS, C. 2012. Quantifying the global burden of morbidity due to unsafe abortion: magnitude in hospital-based studies and methodological issues. *Int J Gynaecol Obstet,* 118 Suppl 2**,** S65-77.

ALEKSANDROWICZ, L., MALHOTRA, V., DIKSHIT, R., GUPTA, P. C., KUMAR, R., SHETH, J., RATHI, S. K., SURAWEERA, W., MIASNIKOF, P., JOTKAR, R., SINHA, D., AWASTHI, S., BHATIA, P. & JHA, P. 2014. Performance criteria for verbal autopsy-based systems to estimate national causes of death: development and application to the Indian Million Death Study. *BMC Med,* 12**,** 21.

ALKEMA, L., KANTOROVA, V., MENOZZI, C. & BIDDLECOM, A. 2013. National, regional, and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: a systematic and comprehensive analysis. *Lancet,* 381**,** 1642-52.

ALKEMA, L., RAFTERY, A. E., GERLAND, P., CLARK, S. J., PELLETIER, F., BUETTNER, T. & HEILIG, G. K. 2011. Probabilistic projections of the total fertility rate for all countries. *Demography,* 48**,** 815-39.

AMOUZOU, A., BANDA, B., KACHAKA, W., JOOS, O., KANYUKA, M., HILL, K. & BRYCE, J. 2014. Monitoring child mortality through community health worker reporting of births and deaths in Malawi: validation against a household mortality survey. *PLoS One,* 9**,** e88939.

AMOUZOU, A., KACHAKA, W., BANDA, B., CHIMZIMU, M., HILL, K. & BRYCE, J. 2013. Monitoring child survival in 'real time' using routine health facility records: results from Malawi. *Trop Med Int Health,* 18**,** 1231-9.

ARAZ, O. M., BENTLEY, D. & MUELLEMAN, R. L. 2014. Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska. *Am J Emerg Med,* 32**,** 1016-23.

ASNAKE, M., HENRY, E. G., TILAHUN, Y. & OLIVERAS, E. 2013. Addressing unmet need for long-acting family planning in Ethiopia: uptake of single-rod progestogen contraceptive implants (Implanon) and characteristics of users. *Int J Gynaecol Obstet,* 123 Suppl 1**,** e29-32.

ATUN, R. 2014. Time for a revolution in reporting of global health data. *Lancet,* 384**,** 937-8.

BANGHA, M., DIAGNE, A., BAWAH, A. & SANKOH, O. 2010. Monitoring the millennium development goals: the potential role of the INDEPTH Network. *Glob Health Action,* 3.

BANK/WHO, W. 2014. Global Civil Registration and Vital Statistics Scaling Up Investment Plan, 2015-2024. Washington, DC: World Bank / WHO.

BECKER, S. & DIOP-SIDIBE, N. 2003. Does use of the calendar in surveys reduce heaping? *Stud Fam Plann,* 34**,** 127-32.

BECKER, S. & SOSA, D. 1992. An experiment using a month-by-month calendar in a family planning survey in Costa Rica. *Stud Fam Plann,* 23**,** 386-91.

BELLI, R. F. 1998. The structure of autobiographical memory and the event history calendar: potential improvements in the quality of retrospective reports in surveys. *Memory,* 6**,** 383-406.

BHATTI, J. A., RAZZAK, J. A., LAGARDE, E. & SALMI, L. R. 2011. Differences in police, ambulance, and emergency department reporting of traffic injuries on Karachi-Hala road, Pakistan. *BMC Res Notes,* 4**,** 75.

BICEGO, G. T., NKAMBULE, R., PETERSON, I., REED, J., DONNELL, D., GININDZA, H., DUONG, Y. T., PATEL, H., BOCK, N., PHILIP, N., MAO, C. & JUSTMAN, J.

2013. Recent patterns in population-based HIV prevalence in Swaziland. *PLoS One,* 8**,** e77101.

BOERMA, J. T. & WEIR, S. S. 2005. Integrating demographic and epidemiological approaches to research on HIV/AIDS: the proximate-determinants framework. *J Infect Dis,* 191 Suppl 1**,** S61-7.

BONGAARTS, J. 1978. A framework for analyzing the proximate determinants of fertility. *Population and development review***,** 105-132.

BONGAARTS, J. 2014. The impact of family planning programs on unmet need and demand for contraception. *Stud Fam Plann,* 45**,** 247-62.

BONGAARTS, J. & CASTERLINE, J. 2013. Fertility Transition: Is sub-Saharan Africa Different? *Popul Dev Rev,* 38**,** 153-168.

BOR, J., HERBST, A. J., NEWELL, M. L. & BARNIGHAUSEN, T. 2013. Increases in adult life expectancy in rural South Africa: valuing the scale-up of HIV treatment. *Science,* 339**,** 961-5.

BORRELL, L. N. & LANCET, E. A. 2012. Race/ethnicity and all-cause mortality in US adults: revisiting the Hispanic paradox. *Am J Public Health,* 102**,** 836-43.

BOSTEAN, G. 2013. Does selective migration explain the Hispanic paradox? A comparative analysis of Mexicans in the U.S. and Mexico. *J Immigr Minor Health,* 15**,** 624-35.

BOWDEN, S., BRAKER, K., CHECCHI, F. & WONG, S. 2012. Implementation and utilisation of community-based mortality surveillance: a case study from Chad. *Confl Health,* 6**,** 11.

BUCKEE, C. O., WESOLOWSKI, A., EAGLE, N. N., HANSEN, E. & SNOW, R. W. 2013. Mobile phones and malaria: modeling human and parasite travel. *Travel Med Infect Dis,* 11**,** 15-22.

BYASS, P., CALVERT, C., MIIRO-NAKIYINGI, J., LUTALO, T., MICHAEL, D., CRAMPIN, A., GREGSON, S., TAKARUZA, A., ROBERTSON, L., HERBST, K., TODD, J. & ZABA, B. 2013. InterVA-4 as a public health tool for measuring HIV/AIDS mortality: a validation study from five African countries. *Glob Health Action,* 6**,** 22448.

BYASS, P., CHANDRAMOHAN, D., CLARK, S. J., D'AMBRUOSO, L., FOTTRELL, E., GRAHAM, W. J., HERBST, A. J., HODGSON, A., HOUNTON, S., KAHN, K., KRISHNAN, A., LEITAO, J., ODHIAMBO, F., SANKOH, O. A. & TOLLMAN, S.

M. 2012. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Glob Health Action,* 5**,** 1-8.

CALEO, G. M., SY, A. P., BALANDINE, S., POLONSKY, J., PALMA, P. P., GRAIS, R. F. & CHECCHI, F. 2012. Sentinel site community surveillance of mortality and nutritional status in southwestern Central African Republic, 2010. *Popul Health Metr,* 10**,** 18.

CALLAHAN, R. L. & BECKER, S. 2012. The reliability of calendar data for reporting contraceptive use: evidence from rural Bangladesh. *Stud Fam Plann,* 43**,** 213-22.

CALVERT, C. & RONSMANS, C. 2013. The contribution of HIV to pregnancy-related mortality: a systematic review and meta-analysis. *AIDS,* 27**,** 1631-9.

CASTERLINE, J. B. & EL-ZEINI, L. O. 2007. The estimation of unwanted fertility. *Demography,* 44**,** 729-45.

CASTERLINE, J. B., PEREZ, A. E. & BIDDLECOM, A. E. 1997. Factors underlying unmet need for family planning in the Philippines. *Stud Fam Plann,* 28**,** 173-91.

CHANDRAMOHAN, D., RODRIGUES, L. C., MAUDE, G. H. & HAYES, R. J. 1998. The validity of verbal autopsies for assessing the causes of institutional maternal death. *Stud Fam Plann,* 29**,** 414-22.

CHEVAN, A. & SUTHERLAND, M. 2009. Revisiting Das Gupta: refinement and extension of standardization and decomposition. *Demography,* 46**,** 429-49.

CHIDAMBARAM, V. C. & PULLUM, T. W. 1981. Estimating fertility trends from retrospective birth histories: Sensitivity to imputation of missing dates. *Popul Stud (Camb),* 35**,** 307-20.

CHIMERE-DAN, O. 1990. Proximate determinants of fertility in Nigeria. *Soc Biol,* 37**,** 162-71.

CLELAND, J., BOERMA, J. T., CARAEL, M. & WEIR, S. S. 2004. Monitoring sexual behaviour in general populations: a synthesis of lessons of the past decade. *Sex Transm Infect,* 80 Suppl 2**,** ii1-7.

CORSI, D. J., NEUMAN, M., FINLAY, J. E. & SUBRAMANIAN, S. V. 2012. Demographic and health surveys: a profile. *Int J Epidemiol,* 41**,** 1602-13.

D'AMICO, M., AGOZZINO, E., BIAGINO, A., SIMONETTI, A. & MARINELLI, P. 1999. Ill-defined and multiple causes on death certificates--a study of misclassification in mortality statistics. *Eur J Epidemiol,* 15**,** 141-8.

DANEL, I. & BORTMAN, M. 2009. An Assessment of LAC's Vital Statistics System: The Foundation of Maternal and Infant Mortality Monitoring. *In:* BANK, W. (ed.). Washington, D.C.: World Bank.

DELAUNAY, V., DOUILLOT, L., DIALLO, A., DIONE, D., TRAPE, J. F., MEDIANIKOV, O., RAOULT, D. & SOKHNA, C. 2013. Profile: the Niakhar Health and Demographic Surveillance System. *Int J Epidemiol,* 42**,** 1002-11.

DELAVANDE, A. & KOHLER, H. P. 2009. Subjective expectations in the context of HIV/AIDS in Malawi. *Demogr Res,* 20**,** 817-874.

DELAVANDE, A. & KOHLER, H. P. 2012. The impact of HIV testing on subjective expectations and risky behavior in Malawi. *Demography,* 49**,** 1011-36.

DELAVANDE, A. & MANSKI, C. F. 2012. Candidate preferences and expectations of election outcomes. *Proc Natl Acad Sci U S A,* 109**,** 3711-5.

DESESQUELLES, A., DEMURU, E., SALVATORE, M. A., PAPPAGALLO, M., FROVA, L., MESLE, F. & EGIDI, V. 2014. Mortality from Alzheimer's disease, Parkinson's disease, and dementias in France and Italy: a comparison using the multiple cause-of-death approach. *J Aging Health,* 26**,** 283-315.

DORRINGTON, R., BOURNE, D., BRADSHAW, D., LAUBSCHER, R. & TIMAEUS, I. M. 2002. HIV/AIDS data in South Africa. *Lancet,* 360**,** 1177.

DORRINGTON, R., BRADSHAW, D., BOURNE, D. & KARIM, S. A. 2000. HIV surveillance results--little grounds for optimism yet. *S Afr Med J,* 90**,** 452-3.

DRUMMOND, M. B. 2011. The Hispanic paradox unraveled? *Am J Respir Crit Care Med,* 184**,** 1222-3.

DUGAS, A. F., JALALPOUR, M., GEL, Y., LEVIN, S., TORCASO, F., IGUSA, T. & ROTHMAN, R. E. 2013. Influenza forecasting with Google Flu Trends. *PLoS One,* 8**,** e56176.

EL ARIFEEN, S., HILL, K., AHSAN, K. Z., JAMIL, K., NAHAR, Q. & STREATFIELD, P. K. 2014. Maternal mortality in Bangladesh: a Countdown to 2015 country case study. *Lancet,* 384**,** 1366-74.

FEDELI, U., SCHIEVANO, E., SAUGO, M. & RODEGHIERO, F. 2014. Mortality from myelodysplastic syndromes: a multiple causes of death approach. *Am J Hematol,* 89**,** 450-1.

FINK, A. K., GERMAN, R. R., HERON, M., STEWART, S. L., JOHNSON, C. J., FINCH, J. L., YIN, D., SCHAEFFER, P. E. & ACCURACY OF CANCER MORTALITY WORKING, G. 2012. Impact of using multiple causes of death codes to compute

site-specific, death certificate-based cancer mortality statistics in the United States. *Cancer Epidemiol,* 36**,** 22-8.

FOTTRELL, E., KAHN, K., TOLLMAN, S. & BYASS, P. 2011. Probabilistic methods for verbal autopsy interpretation: InterVA robustness in relation to variations in a priori probabilities. *PLoS One,* 6**,** e27200.

FRANZINI, L., RIBBLE, J. C. & KEDDIE, A. M. 2001. Understanding the Hispanic paradox. *Ethn Dis,* 11**,** 496-518.

FREEDMAN, D., THORNTON, A., CAMBURN, D., ALWIN, D. & YOUNG-DEMARCO, L. 1988. The life history calendar: a technique for collecting retrospective data. *Sociol Methodol,* 18**,** 37-68.

GAKIDOU, E., HOGAN, M. & LOPEZ, A. D. 2004. Adult mortality: time for a reappraisal. *Int J Epidemiol,* 33**,** 710-7.

GAKIDOU, E. & KING, G. 2006. Death by survey: estimating adult mortality without selection bias from sibling survival data. *Demography,* 43**,** 569-85.

GANATRA, B., TUNCALP, O., JOHNSTON, H. B., JOHNSON, B. R., JR., GULMEZOGLU, A. M. & TEMMERMAN, M. 2014. From concept to measurement: operationalizing WHO's definition of unsafe abortion. *Bull World Health Organ,* 92**,** 155.

GERDTS, C., VOHRA, D. & AHERN, J. 2013. Measuring unsafe abortion-related mortality: a systematic review of the existing methods. *PLoS One,* 8**,** e53346.

GERLAND, P., RAFTERY, A. E., SEVCIKOVA, H., LI, N., GU, D., SPOORENBERG, T., ALKEMA, L., FOSDICK, B. K., CHUNN, J., LALIC, N., BAY, G., BUETTNER, T., HEILIG, G. K. & WILMOTH, J. 2014. World population stabilization unlikely this century. *Science,* 346**,** 234-7.

GUPTA, P. D. 1991. Decomposition of the difference between two rates and its consistency when more than two populations are involved. *Mathematical Population Studies,* 3**,** 105-125.

GUPTA, P. D. & CENSUS, U. S. B. O. T. 1993. *Standardization and decomposition of rates: A user's manual*, US Department of Commerce, Economics and Statistics Administration, Bureau of the Census.

HALLETT, T. B., GREGSON, S., KURWA, F., GARNETT, G. P., DUBE, S., CHAWIRA, G., MASON, P. R. & NYAMUKAPA, C. A. 2010. Measuring and correcting biased child mortality statistics in countries with generalized epidemics of HIV infection. *Bull World Health Organ,* 88**,** 761-8.

HARTTGEN, K., KOWAL, P., STRULIK, H., CHATTERJI, S. & VOLLMER, S. 2013. Patterns of frailty in older adults: comparing results from higher and lower income countries using the Survey of Health, Ageing and Retirement in Europe (SHARE) and the Study on Global AGEing and Adult Health (SAGE). *PLoS One,* 8**,** e75847.

HAYFORD, S. R., AGADJANIAN, V. & LUZ, L. 2012. Now or never: perceived HIV status and fertility intentions in rural Mozambique. *Stud Fam Plann,* 43**,** 191-9.

HAYFORD, S. R. & MORGAN, S. P. 2008. Religiosity and Fertility in the United States: The Role of Fertility Intentions. *Soc Forces,* 86**,** 1163-1188.

HELLERINGER, S., DUTHE, G., KANTE, A. M., ANDRO, A., SOKHNA, C., TRAPE, J. F. & PISON, G. 2013. Misclassification of pregnancy-related deaths in adult mortality surveys: case study in Senegal. *Trop Med Int Health,* 18**,** 27-34.

HELLERINGER, S., KOHLER, H. P., KALILANI-PHIRI, L., MKANDAWIRE, J. & ARMBRUSTER, B. 2011. The reliability of sexual partnership histories: implications for the measurement of partnership concurrency during surveys. *AIDS,* 25**,** 503-11.

HELLERINGER, S., PISON, G., KANTE, A. M., DUTHE, G. & ANDRO, A. 2014a. Reporting errors in siblings' survival histories and their impact on adult mortality estimates: results from a record linkage study in Senegal. *Demography,* 51**,** 387-411.

HELLERINGER, S., PISON, G., MASQUELIER, B., KANTE, A. M., DOUILLOT, L., DUTHE, G., SOKHNA, C. & DELAUNAY, V. 2014b. Improving the quality of adult mortality data collected in demographic surveys: validation study of a new siblings' survival questionnaire in Niakhar, Senegal. *PLoS Med,* 11**,** e1001652.

HEWETT, P. C., MENSCH, B. S., RIBEIRO, M. C., JONES, H. E., LIPPMAN, S. A., MONTGOMERY, M. R. & VAN DE WIJGERT, J. H. 2008. Using sexually transmitted infection biomarkers to validate reporting of sexual behavior within a randomized, experimental evaluation of interviewing methods. *Am J Epidemiol,* 168**,** 202-11.

HILL, K., CHOI, Y. & TIMÆUS, I. 2005. Unconventional approaches to mortality estimation. *Demographic Research,* 13**,** 281-300.

HILL, K., EL ARIFEEN, S., KOENIG, M., AL-SABIR, A., JAMIL, K. & RAGGERS, H. 2006. How should we measure maternal mortality in the developing world? A

comparison of household deaths and sibling history approaches. *Bull World Health Organ,* 84**,** 173-80.

HILL, K., LOPEZ, A. D., SHIBUYA, K., JHA, P. & MONITORING OF VITAL, E. 2007. Interim measures for meeting needs for health sector data: births, deaths, and causes of death. *Lancet,* 370**,** 1726-35.

HO, A., SHIH, M. & SIMON, P. 2007. Hispanic paradox. *Am J Public Health,* 97**,** 392; author reply 392-3.

HOBCRAFT, J. & LITTLE, R. J. 1984. Fertility exposure analysis: A new method for assessing the contribution of proximate determinants to fertility differentials. *Popul Stud (Camb),* 38**,** 21-45.

HOGAN, M. C., FOREMAN, K. J., NAGHAVI, M., AHN, S. Y., WANG, M., MAKELA, S. M., LOPEZ, A. D., LOZANO, R. & MURRAY, C. J. 2010. Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. *Lancet,* 375**,** 1609-23.

JHA, P. 2014. Reliable direct measurement of causes of death in low- and middle-income countries. *BMC Med,* 12**,** 19.

JHA, P., GAJALAKSHMI, V., GUPTA, P. C., KUMAR, R., MONY, P., DHINGRA, N., PETO, R. & COLLABORATORS, R.-C. P. S. 2006. Prospective study of one million deaths in India: rationale, design, and validation results. *PLoS Med,* 3**,** e18.

JHA, P., JACOB, B. & KUMAR, R. 2007. Commentary: Reliable measurement of the causes of mortality in developing countries. *Int J Epidemiol,* 36**,** 651-3.

JUSTMAN, J. E., BENNING, L., DANOFF, A., MINKOFF, H., LEVINE, A., GREENBLATT, R. M., WEBER, K., PIESSENS, E., ROBISON, E. & ANASTOS, K. 2003. Protease inhibitor use and the incidence of diabetes mellitus in a large cohort of HIV-infected women. *J Acquir Immune Defic Syndr,* 32**,** 298-302.

KABADI, G. S., GEUBBELS, E., LYATUU, I., SMITHSON, P., AMARO, R., MEKU, S., SCHELLENBERG, J. A. & MASANJA, H. 2014. Data Resource Profile: The sentinel panel of districts: Tanzania's national platform for health impact evaluation. *Int J Epidemiol*.

KALULE-SABITI, I. 1984. Bongaarts' proximate determinants of fertility applied to group data from the Kenya Fertility Survey 1977/78. *J Biosoc Sci,* 16**,** 205-18.

KAO, S., CHEN, L. M., SHI, L. & WEINRICH, M. C. 1997. Underreporting and misclassification of maternal mortality in Taiwan. *Acta Obstet Gynecol Scand,* 76**,** 629-36.

KASSEBAUM, N. J., BERTOZZI-VILLA, A., COGGESHALL, M. S., SHACKELFORD, K. A., STEINER, C., HEUTON, K. R., GONZALEZ-MEDINA, D., BARBER, R., HUYNH, C., DICKER, D., TEMPLIN, T., WOLOCK, T. M., OZGOREN, A. A., ABD-ALLAH, F., ABERA, S. F., ABUBAKAR, I., ACHOKI, T., ADELEKAN, A., ADEMI, Z., ADOU, A. K., ADSUAR, J. C., AGARDH, E. E., AKENA, D., ALASFOOR, D., ALEMU, Z. A., ALFONSO-CRISTANCHO, R., ALHABIB, S., ALI, R., AL KAHBOURI, M. J., ALLA, F., ALLEN, P. J., ALMAZROA, M. A., ALSHARIF, U., ALVAREZ, E., ALVIS-GUZMAN, N., AMANKWAA, A. A., AMARE, A. T., AMINI, H., AMMAR, W., ANTONIO, C. A., ANWARI, P., ARNLOV, J., ARSENIJEVIC, V. S., ARTAMAN, A., ASAD, M. M., ASGHAR, R. J., ASSADI, R., ATKINS, L. S., BADAWI, A., BALAKRISHNAN, K., BASU, A., BASU, S., BEARDSLEY, J., BEDI, N., BEKELE, T., BELL, M. L., BERNABE, E., BEYENE, T. J., BHUTTA, Z., BIN ABDULHAK, A., BLORE, J. D., BASARA, B. B., BOSE, D., BREITBORDE, N., CARDENAS, R., CASTANEDA-ORJUELA, C. A., CASTRO, R. E., CATALA-LOPEZ, F., CAVLIN, A., CHANG, J. C., CHE, X., CHRISTOPHI, C. A., CHUGH, S. S., CIRILLO, M., COLQUHOUN, S. M., COOPER, L. T., COOPER, C., DA COSTA LEITE, I., DANDONA, L., DANDONA, R., DAVIS, A., DAYAMA, A., DEGENHARDT, L., DE LEO, D., DEL POZO-CRUZ, B., DERIBE, K., DESSALEGN, M., DEVEBER, G. A., DHARMARATNE, S. D., DILMEN, U., DING, E. L., DORRINGTON, R. E., DRISCOLL, T. R., ERMAKOV, S. P., ESTEGHAMATI, A., FARAON, E. J., FARZADFAR, F., FELICIO, M. M., FERESHTEHNEJAD, S. M., DE LIMA, G. M., et al. 2014. Global, regional, and national levels and causes of maternal mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet,* 384**,** 980-1004.

KERBER, K. J., LAWN, J. E., JOHNSON, L. F., MAHY, M., DORRINGTON, R. E., PHILLIPS, H., BRADSHAW, D., NANNAN, N., MSEMBURI, W., OESTERGAARD, M. Z., WALKER, N. P., SANDERS, D. & JACKSON, D. 2013. South African child deaths 1990-2011: have HIV services reversed the trend enough to meet Millennium Development Goal 4? *AIDS,* 27**,** 2637-48.

KHAN, M. F. & SHIRMEEN, A. 2007. Proximate determinants of fertility and reproductive health. *Ulster Med J,* 76**,** 6-7.

KING, G. & LU, Y. 2008. Verbal autopsy methods with multiple causes of death. *Statistical Science,* 23**,** 78-91.

KODZI, I. A., CASTERLINE, J. B. & AGLOBITSE, P. 2010a. The time dynamics of individual fertility preferences among rural Ghanaian women. *Stud Fam Plann,* 41**,** 45-54.

KODZI, I. A., JOHNSON, D. R. & CASTERLINE, J. B. 2010b. Examining the predictive value of fertility preferences among Ghanaian women. *Demogr Res,* 22**,** 965-984.

KOWAL, P., CHATTERJI, S., NAIDOO, N., BIRITWUM, R., FAN, W., LOPEZ RIDAURA, R., MAXIMOVA, T., AROKIASAMY, P., PHASWANA-MAFUYA, N., WILLIAMS, S., SNODGRASS, J. J., MINICUCI, N., D'ESTE, C., PELTZER, K., BOERMA, J. T. & COLLABORATORS, S. 2012. Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). *Int J Epidemiol,* 41**,** 1639-49.

LAGARDE, E. 2007. Road traffic injury is an escalating burden in Africa and deserves proportionate research efforts. *PLoS Med,* 4**,** e170.

LAZER, D., KENNEDY, R., KING, G. & VESPIGNANI, A. 2014. Big data. The parable of Google Flu: traps in big data analysis. *Science,* 343**,** 1203-5.

LE COEUR, S., KHLAT, M., HALEMBOKAKA, G., AUGEREAU-VACHER, C., BATALA-M'PONDO, G., BATY, G. & RONSMANS, C. 2005. HIV and the magnitude of pregnancy-related mortality in Pointe Noire, Congo. *AIDS,* 19**,** 69-75.

LEDDIN, D., ARMSTRONG, D., BORGAONKAR, M., BRIDGES, R. J., FALLONE, C. A., TELFORD, J. J., CHEN, Y., COLACINO, P. & SINCLAIR, P. 2013. The 2012 SAGE wait times program: Survey of Access to GastroEnterology in Canada. *Can J Gastroenterol,* 27**,** 83-9.

LOZANO, R., WANG, H., FOREMAN, K. J., RAJARATNAM, J. K., NAGHAVI, M., MARCUS, J. R., DWYER-LINDGREN, L., LOFGREN, K. T., PHILLIPS, D., ATKINSON, C., LOPEZ, A. D. & MURRAY, C. J. 2011. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis. *Lancet,* 378**,** 1139-65.

MAHAPATRA, P., SHIBUYA, K., LOPEZ, A. D., COULLARE, F., NOTZON, F. C., RAO, C. & SZRETER, S. 2007. Civil registration systems and vital statistics: successes and missed opportunities. *The Lancet,* 370**,** 1653-1663.

MANDL, K. D. 2014. Ebola in the United States: EHRs as a public health tool at the point of care. *JAMA*.

MASQUELIER, B. 2013. Adult mortality from sibling survival data: a reappraisal of selection biases. *Demography,* 50**,** 207-28.

MATHERS, C. D., IBURG, K. M., SALOMON, J. A., TANDON, A., CHATTERJI, S., USTUN, B. & MURRAY, C. J. 2004. Global patterns of healthy life expectancy in the year 2002. *BMC Public Health,* 4**,** 66.

MATHERS, C. D., MURRAY, C. J., LOPEZ, A. D., SADANA, R. & SALOMON, J. A. 2002. Global patterns of healthy life expectancy for older women. *J Women Aging,* 14**,** 99-117.

MATHERS, C. D., SADANA, R., SALOMON, J. A., MURRAY, C. J. & LOPEZ, A. D. 2001. Healthy life expectancy in 191 countries, 1999. *Lancet,* 357**,** 1685-91.

MCCOY, S. I., BUZDUGAN, R., RALPH, L. J., MUSHAVI, A., MAHOMVA, A., HAKOBYAN, A., WATADZAUSHE, C., DIRAWO, J., COWAN, F. M. & PADIAN, N. S. 2014. Unmet need for family planning, contraceptive failure, and unintended pregnancy among HIV-infected and HIV-uninfected women in Zimbabwe. *PLoS One,* 9**,** e105320.

MENSCH, B. S., HEWETT, P. C. & ERULKAR, A. S. 2003. The reporting of sensitive behavior by adolescents: a methodological experiment in Kenya. *Demography,* 40**,** 247-68.

MENSCH, B. S., HEWETT, P. C., GREGORY, R. & HELLERINGER, S. 2008a. Sexual behavior and STI/HIV status among adolescents in rural Malawi: an evaluation of the effect of interview mode on reporting. *Stud Fam Plann,* 39**,** 321-34.

MENSCH, B. S., HEWETT, P. C., JONES, H. E., LUPPI, C. G., LIPPMAN, S. A., PINHO, A. A. & DIAZ, J. 2008b. Consistency in women's reports of sensitive behavior in an interview mode experiment, Sao Paulo, Brazil. *Int Fam Plan Perspect,* 34**,** 169-76.

MIDHET, F. 2008. Validating the verbal autopsy questionnaire for maternal mortality in pakistan. *Int J Health Sci (Qassim),* 2**,** 91-6.

MILINOVICH, G. J., MAGALHÃES, R. J. S., HU, W., O'DONOVAN, J., BERSIN, A., WAGMAN, J. A., GRAY, R. H., CAMPBELL, J. C., THOMA, M. & NDYANABO,

A. 2014. Role of big data in the early detection of Ebola and other emerging infectious diseases. *The Lancet Global Health*.

MILLION DEATH STUDY, C., BASSANI, D. G., KUMAR, R., AWASTHI, S., MORRIS, S. K., PAUL, V. K., SHET, A., RAM, U., GAFFEY, M. F., BLACK, R. E. & JHA, P. 2010. Causes of neonatal and child mortality in India: a nationally representative mortality survey. *Lancet,* 376**,** 1853-60.

MISGANAW, A., MARIAM, D. H., ARAYA, T. & ANENEH, A. 2012. Validity of verbal autopsy method to determine causes of death among adults in the urban setting of Ethiopia. *BMC Med Res Methodol,* 12**,** 130.

MITRA, S. 2013. A data revolution for disability-inclusive development. *Lancet Glob Health,* 1**,** e178-9.

MOBLEY, C. C., BOERMA, J. T., TITUS, S., LOHRKE, B., SHANGULA, K. & BLACK, R. E. 1996. Validation study of a verbal autopsy method for causes of childhood mortality in Namibia. *J Trop Pediatr,* 42**,** 365-9.

MOSSONG, J., BYASS, P. & HERBST, K. 2014. Who died of what in rural KwaZulu-Natal, South Africa: a cause of death analysis using InterVA-4. *Glob Health Action,* 7**,** 25496.

MOULTRIE, T., DORRINGTON, R. E., HILL, A. G., TIMAEUS, I. M. & ZABA, B. 2013. *Tools for demographic estimation,* Paris, International Union for the Scientific Study of Population.

MOULTRIE, T. A., HOSEGOOD, V., MCGRATH, N., HILL, C., HERBST, K. & NEWELL, M. L. 2008. Refining the criteria for stalled fertility declines: an application to rural KwaZulu-Natal, South Africa, 1990-2005. *Stud Fam Plann,* 39**,** 39-48.

MOULTRIE, T. A., SAYI, T. S. & TIMAEUS, I. M. 2012. Birth intervals, postponement, and fertility decline in Africa: a new type of transition? *Popul Stud (Camb),* 66**,** 241-58.

MURRAY, C. J., LOPEZ, A. D., FEEHAN, D. M., PETER, S. T. & YANG, G. 2007. Validation of the symptom pattern method for analyzing verbal autopsy data. *PLoS Med,* 4**,** e327.

NAGHAVI, M., MAKELA, S., FOREMAN, K., O'BRIEN, J., POURMALEK, F. & LOZANO, R. 2010. Algorithms for enhancing public health utility of national causes-of-death data. *Popul Health Metr,* 8**,** 9.

NDILA, C., BAUNI, E., MOCHAMAH, G., NYIRONGO, V., MAKAZI, A., KOSGEI, P., TSOFA, B., NYUTU, G., ETYANG, A., BYASS, P. & WILLIAMS, T. N. 2014.

Causes of death among persons of all ages within the Kilifi Health and Demographic Surveillance System, Kenya, determined from verbal autopsies interpreted using the InterVA-4 model. *Glob Health Action,* 7**,** 25593.

NGOM, P., BINKA, F. N., PHILLIPS, J. F., PENCE, B. & MACLEOD, B. 2001. Demographic surveillance and health equity in sub-Saharan Africa. *Health Policy Plan,* 16**,** 337-44.

NOVE, A., MATTHEWS, Z., NEAL, S. & CAMACHO, A. V. 2014. Maternal mortality in adolescents compared with women of other ages: evidence from 144 countries. *Lancet Glob Health,* 2**,** e155-64.

OBERMEYER, Z., RAJARATNAM, J. K., PARK, C. H., GAKIDOU, E., HOGAN, M. C., LOPEZ, A. D. & MURRAY, C. J. 2010. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006. *PLoS Med,* 7**,** e1000260.

ODURO, A. R., WAK, G., AZONGO, D., DEBPUUR, C., WONTUO, P., KONDAYIRE, F., WELAGA, P., BAWAH, A., NAZZAR, A., WILLIAMS, J., HODGSON, A. & BINKA, F. 2012. Profile of the Navrongo Health and Demographic Surveillance System. *Int J Epidemiol,* 41**,** 968-76.

PACHECO, A. G., SARACENI, V., TUBOI, S. H., LAURIA, L. M., MOULTON, L. H., FAULHABER, J. C., KING, B., GOLUB, J. E., DUROVNI, B., CAVALCANTE, S., HARRISON, L. H., CHAISSON, R. E. & SCHECHTER, M. 2011. Estimating the extent of underreporting of mortality among HIV-infected individuals in Rio de Janeiro, Brazil. *AIDS Res Hum Retroviruses,* 27**,** 25-8.

PALLONI, A. & ARIAS, E. 2004. Paradox lost: explaining the Hispanic adult mortality advantage. *Demography,* 41**,** 385-415.

PALLONI, A. & MORENOFF, J. D. 2001. Interpreting the paradoxical in the hispanic paradox: demographic and epidemiologic approaches. *Ann N Y Acad Sci,* 954**,** 140-74.

PARRADO, E. A. 2011. How high is Hispanic/Mexican fertility in the united states? Immigration and tempo considerations. *Demography,* 48**,** 1059-80.

PATEL, K. V., ESCHBACH, K., RAY, L. A. & MARKIDES, K. S. 2004. Evaluation of mortality data for older Mexican Americans: implications for the Hispanic paradox. *Am J Epidemiol,* 159**,** 707-15.

PAYNE, C. F., MKANDAWIRE, J. & KOHLER, H. P. 2013. Disability transitions and health expectancies among adults 45 years and older in Malawi: a cohort-based model. *PLoS Med,* 10**,** e1001435.

PEARSON, E. & BECKER, S. 2014. Couples' unmet need for family planning in three west african countries. *Stud Fam Plann,* 45**,** 339-59.

PERVAIZ, F., PERVAIZ, M., ABDUR REHMAN, N. & SAIF, U. 2012. FluBreaks: early epidemic detection from Google flu trends. *J Med Internet Res,* 14**,** e125.

PHILLIPS, D. E., LOZANO, R., NAGHAVI, M., ATKINSON, C., GONZALEZ-MEDINA, D., MIKKELSEN, L., MURRAY, C. J. & LOPEZ, A. D. 2014. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Popul Health Metr,* 12**,** 14.

PINHEIRO, P. S., WILLIAMS, M., MILLER, E. A., EASTERDAY, S., MOONIE, S. & TRAPIDO, E. J. 2011. Cancer survival among Latinos and the Hispanic Paradox. *Cancer Causes Control,* 22**,** 553-61.

PISON, G., DOUILLOT, L., KANTE, A. M., NDIAYE, O., DIOUF, P. N., SENGHOR, P., SOKHNA, C. & DELAUNAY, V. 2014. Health & demographic surveillance system profile: Bandafassi Health and Demographic Surveillance System (Bandafassi HDSS), Senegal. *Int J Epidemiol,* 43**,** 739-48.

PRESTON, S. H. & ELO, I. T. 2014. Anatomy of a Municipal Triumph: New York City's Upsurge in Life Expectancy. *Population and Development Review,* 40**,** 1-29.

PRESTON, S. H., ELO, I. T., ROSENWAIKE, I. & HILL, M. 1996. African-American mortality at older ages: results of a matching study. *Demography,* 33**,** 193-209.

PRESTON, S. H., HEUVELINE, P. & GUILLOT, M. 2001. *Demography : measuring and modeling population processes,* Malden, MA, Blackwell Publishers.

PULLUM, T. W. 1991. Statistical methods to adjust for date and age misreporting to improve estimates of vital rates in Pakistan. *Stat Med,* 10**,** 191-200.

PULLUM, T. W., SCHOUMAKER, B., BECKER, S. & BRADLEY, S. E. An assessment of DHS estimates of fertility and under-five mortality. XXVII International Population Conference of the International Union for the Scientific Study of Population, Busan, Korea, 2013. 26-31.

QUIGLEY, M. A., CHANDRAMOHAN, D., SETEL, P., BINKA, F. & RODRIGUES, L. C. 2000. Validity of data-derived algorithms for ascertaining causes of adult death in two African sites using verbal autopsy. *Trop Med Int Health,* 5**,** 33-9.

RAFTERY, A. E., ALKEMA, L. & GERLAND, P. 2014a. Bayesian Population Projections for the United Nations. *Stat Sci,* 29**,** 58-68.

RAFTERY, A. E., CHUNN, J. L., GERLAND, P. & SEVCIKOVA, H. 2013. Bayesian probabilistic projections of life expectancy for all countries. *Demography,* 50**,** 777-801.

RAFTERY, A. E., LALIC, N. & GERLAND, P. 2014b. Joint Probabilistic Projection of Female and Male Life Expectancy. *Demogr Res,* 30**,** 795-822.

RAFTERY, A. E., LI, N., SEVCIKOVA, H., GERLAND, P. & HEILIG, G. K. 2012. Bayesian probabilistic population projections for all countries. *Proc Natl Acad Sci U S A,* 109**,** 13915-21.

RAMROTH, H., LORENZ, E., RANKIN, J. C., FOTTRELL, E., YE, M., NEUHANN, F., SSENNONO, M., SIE, A., BYASS, P. & BECHER, H. 2012. Cause of death distribution with InterVA and physician coding in a rural area of Burkina Faso. *Trop Med Int Health,* 17**,** 904-13.

RENIERS, G., SLAYMAKER, E., NAKIYINGI-MIIRO, J., NYAMUKAPA, C., CRAMPIN, A. C., HERBST, K., URASSA, M., OTIENO, F., GREGSON, S., SEWE, M., MICHAEL, D., LUTALO, T., HOSEGOOD, V., KASAMBA, I., PRICE, A., NABUKALU, D., MCLEAN, E., ZABA, B. & NETWORK, A. 2014. Mortality trends in the era of antiretroviral therapy: evidence from the Network for Analysing Longitudinal Population based HIV/AIDS data on Africa (ALPHA). *AIDS,* 28 Suppl 4**,** S533-42.

ROBERTS, B., MORGAN, O. W., SULTANI, M. G., NYASULU, P., RWEBANGILA, S., MYATT, M., SONDORP, E., CHANDRAMOHAN, D. & CHECCHI, F. 2010. A new method to estimate mortality in crisis-affected and resource-poor settings: validation study. *Int J Epidemiol,* 39**,** 1584-96.

ROBERTS, B., MORGAN, O. W., SULTANI, M. G., NYASULU, P., RWEBANGILA, S., SONDORP, E., CHANDRAMOHAN, D. & CHECCHI, F. 2011. Economic feasibility of a new method to estimate mortality in crisis-affected and resource-poor settings. *PLoS One,* 6**,** e25175.

ROSSIER, C., MUINDI, K., SOURA, A., MBERU, B., LANKOANDE, B., KABIRU, C. & MILLOGO, R. 2014a. Maternal health care utilization in Nairobi and Ouagadougou: evidence from HDSS. *Glob Health Action,* 7**,** 24351.

ROSSIER, C., SOURA, A., BAYA, B., COMPAORE, G., DABIRE, B., DOS SANTOS, S., DUTHE, G., GNOUMOU, B., KOBIANE, J. F., KOUANDA, S., LANKOANDE, B.,

LEGRAND, T., MBACKE, C., MILLOGO, R., MONDAIN, N., MONTGOMERY, M., NIKIEMA, A., OUILI, I., PISON, G., RANDALL, S., SANGLI, G., SCHOUMAKER, B. & ZOURKALEINI, Y. 2012. Profile: the Ouagadougou Health and Demographic Surveillance System. *Int J Epidemiol,* 41**,** 658-66.

ROSSIER, C., SOURA, A. B., DUTHE, G. & FINDLEY, S. 2014b. Non-Communicable Disease Mortality and Risk Factors in Formal and Informal Neighborhoods, Ouagadougou, Burkina Faso: Evidence from a Health and Demographic Surveillance System. *PLoS One,* 9**,** e113780.

RUGG, D., MARAIS, H., CARAEL, M., DE LAY, P. & WARNER-SMITH, M. 2009. Are we on course for reporting on the Millennium Development Goals in 2015? *J Acquir Immune Defic Syndr,* 52 Suppl 2**,** S69-76.

SACHS, J. D. 2012. From millennium development goals to sustainable development goals. *Lancet,* 379**,** 2206-11.

SANKOH, O. & BYASS, P. 2014. Time for civil registration with verbal autopsy. *The Lancet Global Health,* 2**,** e693-e694.

SANKOH, O. A., NGOM, P., CLARK, S. J., DE SAVIGNY, D. & BINKA, F. 2006. Levels and Patterns of Mortality at INDEPTH Demographic Surveillance Systems. *In:* JAMISON, D. T., FEACHEM, R. G., MAKGOBA, M. W., BOS, E. R., BAINGANA, F. K., HOFMAN, K. J. & ROGO, K. O. (eds.) *Disease and Mortality in Sub-Saharan Africa.* 2nd ed. Washington (DC).

SANTILLANA, M., ZHANG, D. W., ALTHOUSE, B. M. & AYERS, J. W. 2014. What can digital disease detection learn from (an external revision to) google flu trends? *Am J Prev Med,* 47**,** 341-7.

SDSN 2015. Indicators and a monitoring framework for Sustainable Development Goals. Sustainable Development Solutions Network.

SETEL, P. W., MACFARLANE, S. B., SZRETER, S., MIKKELSEN, L., JHA, P., STOUT, S., ABOUZAHR, C. & MONITORING OF VITAL, E. 2007. A scandal of invisibility: making everyone count by counting everyone. *Lancet,* 370**,** 1569-77.

SETEL, P. W., WHITING, D. R., HEMED, Y., CHANDRAMOHAN, D., WOLFSON, L. J., ALBERTI, K. G. & LOPEZ, A. D. 2006. Validity of verbal autopsy procedures for determining cause of death in Tanzania. *Trop Med Int Health,* 11**,** 681-96.

SHAH, I. H., PULLUM, T. W. & IRFAN, M. 1986. Fertility in Pakistan during the 1970s. *J Biosoc Sci,* 18**,** 215-29.

SHAHIDULLAH, M. 1995a. A comparison of sisterhood information on causes of maternal death with the registration causes of maternal death in Matlab, Bangladesh. *Int J Epidemiol,* 24**,** 937-42.

SHAHIDULLAH, M. 1995b. The sisterhood method of estimating maternal mortality: the Matlab experience. *Stud Fam Plann,* 26**,** 101-6.

SHAKHATREH, F. M. 2003. Unmet need for family planning. *Saudi Med J,* 24**,** 1268-9.

SHRESTHA, A., STOECKEL, J. & TULADHAR, J. M. 1991. The KAP-gap in Nepal: reasons for non-use of contraception among couples with an unmet need for family planning. *Asia Pac Popul J,* 6**,** 25-38.

SMITH, D. P. & BRADSHAW, B. S. 2006. Rethinking the Hispanic paradox: death rates and life expectancy for US non-Hispanic White and Hispanic populations. *Am J Public Health,* 96**,** 1686-92.

SOBNGWI-TAMBEKOU, J., BHATTI, J., KOUNGA, G., SALMI, L. R. & LAGARDE, E. 2010. Road traffic crashes on the Yaounde-Douala road section, Cameroon. *Accid Anal Prev,* 42**,** 422-6.

SPEIZER, I. S., CALHOUN, L. M., HOKE, T. & SENGUPTA, R. 2013. Measurement of unmet need for family planning: longitudinal analysis of the impact of fertility desires on subsequent childbearing behaviors among urban women from Uttar Pradesh, India. *Contraception,* 88**,** 553-60.

STANTON, C., ABDERRAHIM, N. & HILL, K. 2000. An assessment of DHS maternal mortality indicators. *Stud Fam Plann,* 31**,** 111-23.

STATE, B., RODRIGUEZ, M., HELBING, D. & ZAGHENI, E. 2014. Migration of Professionals to the U.S.: Evidence from LinkedIn data.

STOLZENBERG, R. M. 2011. Do not go gentle into that good night: the effect of retirement on subsequent mortality of U.S. Supreme Court justices, 1801-2006. *Demography,* 48**,** 1317-46.

STOLZENBERG, R. M. & LINDGREN, J. 2010. Retirement and death in office of U.S. Supreme Court justices. *Demography,* 47**,** 269-98.

STOVER, J. 1998. Revising the proximate determinants of fertility framework: what have we learned in the past 20 years? *Stud Fam Plann,* 29**,** 255-67.

STREATFIELD, P. K., ALAM, N., COMPAORE, Y., ROSSIER, C., SOURA, A. B., BONFOH, B., JAEGER, F., NGORAN, E. K., UTZINGER, J., GOMEZ, P., JASSEH, M., ANSAH, A., DEBPUUR, C., ODURO, A., WILLIAMS, J., ADDEI, S., GYAPONG, M., KUKULA, V. A., BAUNI, E., MOCHAMAH, G., NDILA, C.,

WILLIAMS, T. N., DESAI, M., MOIGE, H., ODHIAMBO, F. O., OGWANG, S., BEGUY, D., EZEH, A., OTI, S., CHIHANA, M., CRAMPIN, A., PRICE, A., DELAUNAY, V., DIALLO, A., DOUILLOT, L., SOKHNA, C., COLLINSON, M. A., KAHN, K., TOLLMAN, S. M., HERBST, K., MOSSONG, J., EMINA, J. B., SANKOH, O. A. & BYASS, P. 2014a. Pregnancy-related mortality in Africa and Asia: evidence from INDEPTH Health and Demographic Surveillance System sites. *Glob Health Action,* 7**,** 25368.

STREATFIELD, P. K., KHAN, W. A., BHUIYA, A., HANIFI, S. M., ALAM, N., DIBOULO, E., SIE, A., YE, M., COMPAORE, Y., SOURA, A. B., BONFOH, B., JAEGER, F., NGORAN, E. K., UTZINGER, J., MELAKU, Y. A., MULUGETA, A., WELDEAREGAWI, B., GOMEZ, P., JASSEH, M., HODGSON, A., ODURO, A., WELAGA, P., WILLIAMS, J., AWINI, E., BINKA, F. N., GYAPONG, M., KANT, S., MISRA, P., SRIVASTAVA, R., CHAUDHARY, B., JUVEKAR, S., WAHAB, A., WILOPO, S., BAUNI, E., MOCHAMAH, G., NDILA, C., WILLIAMS, T. N., HAMEL, M. J., LINDBLADE, K. A., ODHIAMBO, F. O., SLUTSKER, L., EZEH, A., KYOBUTUNGI, C., WAMUKOYA, M., DELAUNAY, V., DIALLO, A., DOUILLOT, L., SOKHNA, C., GOMEZ-OLIVE, F. X., KABUDULA, C. W., MEE, P., HERBST, K., MOSSONG, J., CHUC, N. T., ARTHUR, S. S., SANKOH, O. A., TANNER, M. & BYASS, P. 2014b. Malaria mortality in Africa and Asia: evidence from INDEPTH health and demographic surveillance system sites. *Glob Health Action,* 7**,** 25369.

STREATFIELD, P. K., KHAN, W. A., BHUIYA, A., HANIFI, S. M., ALAM, N., MILLOGO, O., SIE, A., ZABRE, P., ROSSIER, C., SOURA, A. B., BONFOH, B., KONE, S., NGORAN, E. K., UTZINGER, J., ABERA, S. F., MELAKU, Y. A., WELDEAREGAWI, B., GOMEZ, P., JASSEH, M., ANSAH, P., AZONGO, D., KONDAYIRE, F., ODURO, A., AMU, A., GYAPONG, M., KWARTENG, O., KANT, S., PANDAV, C. S., RAI, S. K., JUVEKAR, S., MURALIDHARAN, V., WAHAB, A., WILOPO, S., BAUNI, E., MOCHAMAH, G., NDILA, C., WILLIAMS, T. N., KHAGAYI, S., LASERSON, K. F., NYAGUARA, A., VAN EIJK, A. M., EZEH, A., KYOBUTUNGI, C., WAMUKOYA, M., CHIHANA, M., CRAMPIN, A., PRICE, A., DELAUNAY, V., DIALLO, A., DOUILLOT, L., SOKHNA, C., GOMEZ-OLIVE, F. X., MEE, P., TOLLMAN, S. M., HERBST, K., MOSSONG, J., CHUC, N. T., ARTHUR, S. S., SANKOH, O. A. & BYASS, P. 2014c. HIV/AIDS-related

mortality in Africa and Asia: evidence from INDEPTH health and demographic surveillance system sites. *Glob Health Action,* 7**,** 25370.

TARDON, A. G., ZAPLANA, J., HERNANDEZ, R. & CUETO, A. 1995. Usefulness of the codification of multiple causes of death in mortality statistics. *Int J Epidemiol,* 24**,** 1132-7.

TENSOU, B., ARAYA, T., TELAKE, D. S., BYASS, P., BERHANE, Y., KEBEBEW, T., SANDERS, E. J. & RENIERS, G. 2010. Evaluating the InterVA model for determining AIDS mortality from verbal autopsies in the adult population of Addis Ababa. *Trop Med Int Health,* 15**,** 547-53.

THOMPSON, L. H., MALIK, M. T., GUMEL, A., STROME, T. & MAHMUD, S. M. 2014. Emergency department and 'Google flu trends' data as syndromic surveillance indicators for seasonal influenza. *Epidemiol Infect,* 142**,** 2397-405.

THOMSON, E. F., NURU-JETER, A., RICHARDSON, D., RAZA, F. & MINKLER, M. 2013. The Hispanic Paradox and older adults' disabilities: is there a healthy migrant effect? *Int J Environ Res Public Health,* 10**,** 1786-814.

TIEN, P. C., SCHNEIDER, M. F., COLE, S. R., LEVINE, A. M., COHEN, M., DEHOVITZ, J., YOUNG, M. & JUSTMAN, J. E. 2007. Antiretroviral therapy exposure and incidence of diabetes mellitus in the Women's Interagency HIV Study. *AIDS,* 21**,** 1739-45.

TRAPE, J. F., SAUVAGE, C., NDIAYE, O., DOUILLOT, L., MARRA, A., DIALLO, A., CISSE, B., GREENWOOD, B., MILLIGAN, P., SOKHNA, C. & MOLEZ, J. F. 2012. New malaria-control policies and child mortality in senegal: reaching millennium development goal 4. *J Infect Dis,* 205**,** 672-9.

TRINITAPOLI, J. & YEATMAN, S. 2011. Uncertainty and Fertility in a Generalized AIDS Epidemic. *Am Sociol Rev,* 76**,** 935-954.

TURRA, C. M. & GOLDMAN, N. 2007. Socioeconomic differences in mortality among U.S. adults: insights into the Hispanic paradox. *J Gerontol B Psychol Sci Soc Sci,* 62**,** S184-92.

UN 1983. *Manual X: indirect techniques for demographic estimation,* New York.

UNITED STATES. AGENCY FOR INTERNATIONAL DEVELOPMENT USAID. CENTER FOR POPULATION, H. & NUTRITION 1998. Unmet need for family planning. *Glob Issues,* 3**,** 24-5.

UNITED-NATIONS 2014. The Millennium Development Goals Report

2014. New York: United Nations.

VERGNANO, S., FOTTRELL, E., OSRIN, D., KAZEMBE, P. N., MWANSAMBO, C., MANANDHAR, D. S., MUNJANJA, S. P., BYASS, P., LEWYCKA, S. & COSTELLO, A. 2011. Adaptation of a probabilistic method (InterVA) of verbal autopsy to improve the interpretation of cause of stillbirth and neonatal death in Malawi, Nepal, and Zimbabwe. *Popul Health Metr,* 9**,** 48.

WANG, H., DWYER-LINDGREN, L., LOFGREN, K. T., RAJARATNAM, J. K., MARCUS, J. R., LEVIN-RECTOR, A., LEVITZ, C. E., LOPEZ, A. D. & MURRAY, C. J. 2012. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2071-94.

WESOLOWSKI, A., BUCKEE, C. O., PINDOLIA, D. K., EAGLE, N., SMITH, D. L., GARCIA, A. J. & TATEM, A. J. 2013. The use of census migration data to approximate human movement patterns across temporal scales. *PLoS One,* 8**,** e52971.

WESOLOWSKI, A., EAGLE, N., TATEM, A. J., SMITH, D. L., NOOR, A. M., SNOW, R. W. & BUCKEE, C. O. 2012. Quantifying the impact of human mobility on malaria. *Science,* 338**,** 267-70.

WESOLOWSKI, A., STRESMAN, G., EAGLE, N., STEVENSON, J., OWAGA, C., MARUBE, E., BOUSEMA, T., DRAKELEY, C., COX, J. & BUCKEE, C. O. 2014. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. *Sci Rep,* 4**,** 5678.

WHITING, D. R., SETEL, P. W., CHANDRAMOHAN, D., WOLFSON, L. J., HEMED, Y. & LOPEZ, A. D. 2006. Estimating cause-specific mortality from community- and facility-based data sources in the United Republic of Tanzania: options and implications for mortality burden estimates. *Bull World Health Organ,* 84**,** 940-8.

WILMOTH, J., MATHERS, C., SAY, L. & MILLS, S. 2010. Maternal deaths drop by one-third from 1990 to 2008: a United Nations analysis. *Bull World Health Organ,* 88**,** 718-718A.

WITHERS, M. H., TAVROW, P. & ADINATA, N. A. 2011. Do ambivalent women have an unmet need for family planning? A longitudinal study from Bali, Indonesia. *Womens Health Issues,* 21**,** 444-9.

WIWANITKIT, V. 2014. Google Flu for forecasting influenza-like illness. *Am J Emerg Med,* 32**,** 1417.

WONG, O., ROCKETTE, H. E., REDMOND, C. K. & HEID, M. 1978. Evaluation of multiple causes of death in occupational mortality studies. *J Chronic Dis,* 31**,** 183-93.

YE, Y., WAMUKOYA, M., EZEH, A., EMINA, J. B. & SANKOH, O. 2012. Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Sahara Africa? *BMC Public Health,* 12**,** 741.

YOUNG, R. P. & HOPKINS, R. J. 2014a. The Hispanic paradox further unraveled? *Thorax,* 69**,** 184-5.

YOUNG, R. P. & HOPKINS, R. J. 2014b. A review of the Hispanic paradox: time to spill the beans? *Eur Respir Rev,* 23**,** 439-49.

ZABA, B., CALVERT, C., MARSTON, M., ISINGO, R., NAKIYINGI-MIIRO, J., LUTALO, T., CRAMPIN, A., ROBERTSON, L., HERBST, K., NEWELL, M. L., TODD, J., BYASS, P., BOERMA, T. & RONSMANS, C. 2013. Effect of HIV infection on pregnancy-related mortality in sub-Saharan Africa: secondary analyses of pooled community-based data from the network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA). *Lancet,* 381**,** 1763-71.

ZAGHENI, E. 2011. The leverage of demographic dynamics on carbon dioxide emissions: does age structure matter? *Demography,* 48**,** 371-99.

ZUREICK-BROWN, S., NEWBY, H., CHOU, D., MIZOGUCHI, N., SAY, L., SUZUKI, E. & WILMOTH, J. 2013. Understanding global trends in maternal mortality. *Int Perspect Sex Reprod Health,* 39**,** 32-41.

ZWANE, A. P., ZINMAN, J., VAN DUSEN, E., PARIENTE, W., NULL, C., MIGUEL, E., KREMER, M., KARLAN, D. S., HORNBECK, R., GINE, X., DUFLO, E., DEVOTO, F., CREPON, B. & BANERJEE, A. 2011. Being surveyed can change later behavior and related parameter estimates. *Proc Natl Acad Sci U S A,* 108**,** 1821-6.