Monitoring Health Equity in the MDGs: A Practical Guide

Meg Wirth
Enrique Delamonica
Emma Sacks
Deborah Balk
Adam Storeygard
Alberto Minujin

CIESIN and UNICEF
January 2006

Queries may be sent to:
Meg Wirth
megwirth@earthlink.net

Table of Contents

Tables

Figures

Boxes

Preface

The global consensus represented by the MDGs is a new point of departure for the development community. The UN Millennium Project and scores of other decision-makers, activists, bilateral aid organizations, and communities are already deeply immersed in efforts to reach the (Sachs et al. 2004, UNDP 2003). Yet this global effort could still benefit from explicit and systematic commitment to equity at the country level. Well-defined, equity-sensitive targets—linked to relevant data sources—are necessary to ensure that poor, marginalized, and vulnerable groups are given opportunities for improved health and access to health services (Freedman et al. 2004, Gwatkin 2003).

The original purpose of this analysis was to show how to monitor the maternal and child health MDGs[1] in an equity-sensitive manner. Using data from broad-scale, international household-level surveys—the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS)—that analysis (which examined six countries using 20 health indicators and six social strata to ground recommendations in current data[2]) demonstrated that in order to understand and promote equity, it is necessary and feasible to establish a baseline using a variety of indicators and stratifiers, even in very low-income, data-poor countries.

This *Practical Guide* builds on the previous work and addresses a wider audience, providing specific technical advice on how to undertake a simple equity analysis using population-based surveys.

---

[1] The Millennium Development Goals (MDGs) are a set of eight poverty alleviation goals set forth by the United Nations at the Millennium Summit in 2000. Goal Four is to reduce child mortality by two thirds and Goal Five is to reduce maternal mortality by three-fourths. All of the goals have a target date of 2015, using a 1990 baseline. Other goals are also relevant to maternal and child health. For example, 'proportion of underweight children' is the indicator used in Goal One ('to halve extreme poverty and hunger').

[2] Analysis and dataset direct link: http://www.ciesin.columbia.edu/pdf/Two_way_Health_Equity_Tables.pdf; http://www.ciesin.columbia.edu/pdf/Three_way_Health_Equity_Tables.pdf; http://www.unmillenniumproject.org/who/tf4docs.htm.

Acknowledgments

The original analysis that provided the framework for this Guide was funded through the UN Millennium Project Funds to Task Force 4 (child health and maternal health). Funding from the World Bank's Japan Policy and Human Resource Development (PHRD) Fund Poverty Mapping Project at CIESIN (The Center for International Earth Science Information Network) at Columbia University's Earth Institute made the development of this document possible. Thank you to Elisabeth Sydor for design work.

Introduction

The health gap between so-called haves and have-nots is a phenomenon documented throughout the world. Health gaps exist between rich and poor, well-educated and less-educated, rural and urban—even between impoverished and marginalized groups. The initiation of the Millennium Development Goals (MDGs) brings new energy and resources to the development agenda, with a strong emphasis on health outcomes. However, ensuring that such goals are met in an equitable manner, with the worst-off groups benefiting at the same or greater rate as the better-off groups, is a central challenge. One place to start is by identifying inequity between different population groups and across different social strata.

Inequity baselines can be created to illuminate how different indicators and social stratifiers yield different patterns of disparity. Using population-based surveys, such baselines can be established even in data-poor countries. Using survey data from Ghana, Ethiopia, Kenya, Tajikistan, Dominican Republic and Cambodia, this *Guide* demonstrates how to establish such a baseline.

This *Guide* is for researchers, program staff, and policymakers interested in undertaking equity-sensitive monitoring of key health (or other development) indicators within a given country (as opposed to cross-national studies). The *Guide* provides a method for analyzing indicators across a number of social strata including wealth, ethnicity, education, region, sex, and geography. The methodology presented includes both single and simultaneous stratification that allow for the generation of fairly simple, quick equity-monitoring tools. The mode of presentation is instructional, rather than theoretical.

Although the examples used here are based on the DHS and MICS, the methods can be easily adapted for use with other surveys, including those that are not nationally representative. Currently, countries lacking vital registration systems rely on population-based surveys for MDG monitoring. As long as a country has at least one population-based household survey containing information on social characteristics as well as health, analysts can also use this data source for equity monitoring

Furthermore, although the *Guide* is based upon the health MDGs, the techniques described herein could be adapted for non-health MDGs (e.g., education, water, and sanitation) as well as other development goals at the national or subnational level.

Policy relevance

These methods can be duplicated in any country or region where there is appropriate data. The type of analysis should be tailored to fit relevant policy issues, either internationally or within a given region. Country-specific considerations, based upon current policy priorities and epidemiological trends, should be taken into account when selecting both indicators and stratifiers.

For example, in choosing indicators in a country where malaria represents a large portion of the disease burden, detailed analysis of malaria indicators should be prioritized. Likewise, both the definition of a skilled birth attendant and the relevance of that definition vary between countries. Other priorities, too, may vary based on vaccination campaign histories and the general epidemiological profile of the country.

In choosing stratifiers, attention must also be given to country-specific considerations. For example, it would not be useful to focus on ethnic disparities as a stratifier in a country where ethnic disparities are minimal or nonexistent. Conversely, in a country or area with extreme ethnic tension, policymakers would need to choose whether to examine ethnicity explicitly or to strategically use geography, for example, as a proxy for race or ethnicity, because of data limitations or the existing political difficulties. The relative importance of occupation and the extent to which regional strata correspond to funding and political/administrative regions may also vary. Lastly, availability of data (e.g., for geo-coordinates or race) in specific surveys will affect which stratifiers and indicators can be used for a given country.

The methods presented here are a starting point, not an end point. In order to be relevant to policymakers, more qualitative work may be needed to complement the quantitative measurement of 'health gaps.' A range of standard behavioral and social science methods must be used to explain and augment the numbers. For example, historical and sociological work on ethnic groupings should inform the analysis and interpretation of this stratifier. Participatory methods could be used to inform policymakers of the health priorities of certain groups and communities. In addition, shortcomings in sampling frames cause vulnerable population groups such as refugee populations, urban slum dwellers, orphans and minorities to be excluded from analyses, as is noted. Additional efforts should be made to augment the analysis to include these groups.

The *Guide* is divided into three parts: Part I covers background information and definitions, including indicator and stratifier selection; Part II includes detailed discussion of *single* and *simultaneous* stratification; and Part III includes information on monitoring, implementation and conclusions.

Abbreviations and Acronyms

AAFM	Age at First Marriage
AIDS	Autoimmune Deficiency Syndrome
CDC	Center for Disease Control and Prevention, U.S.
CPR	Contraceptive Prevalence Rate
CPR-MM	Contraceptive Prevalence Rate—Modern Method
CSHP	Country-Specific Health Professional
DHS	Demographic and Health Surveys
DPT	Diphtheria, Pertussis, and Tetanus Vaccine
DSS	Data Surveillance System
EmOC	Emergency Obstetric Care
GIS	Geographic Information Systems
HDR	Human Development Report
HIV	Human Immunodeficiency Virus
IDB	Inter-American Development Bank
IFLS	Indonesian Family Life Survey
IMR	Infant Mortality Rate
IRHS	International Reproductive Health Surveys
LSMS	Living Standards Measurement Surveys
MDGs	Millennium Development Goals
MICS	Multiple Indicator Cluster Surveys
NNMR	Neonatal Mortality Rate
RNS	Rural with No Schooling
PAPFAM	Pan Arab Project for Family Health
PRSP	Poverty Reduction Strategy Papers
SBA	Skilled Birth Attendant
SE	Standard Error
SES	Socio-economic Status
U5MR	Under-Five Mortality Rate
UE	Urban and Educated
UIS	Urban Inequities Survey
UN	United Nations
UNDP	United Nations Development Programme
UNICEF	United Nations Children’s Fund
USAID	United States Agency for International Development
WDI	World Development Index
WHO	World Health Organization
WHS	World Health Surveys

PART 1
Chapter 1
Who Can Do This Analysis

Although an examination of equity is feasible in even a low-income, resource-poor setting, there is a certain baseline knowledge needed to carry out this process. A team must have enough public health and policy knowledge to choose indicators and stratifiers which are relevant and locally appropriate. Equity analysis must be situated within a clear assessment of the epidemiological and health profile of the population. A study should take into account current demographic, social, and epidemiological patterns and prioritize the most pressing inequity problems.

Depending on the depth desired for the analysis, statistical knowledge and software may be required. This *Guide* is based on analysis using SPSS, as well as Microsoft Excel. STATA was used for calculating significance for simultaneously-stratified data; Microsoft Excel was used to make graphs and charts; and ArcGIS was used for mapping.

The human resources required for this project may vary based upon the breadth of the indicators and stratifiers under study. This analysis does not require advanced statistical abilities, or even multivariate regression analysis. Because many different data files and formats are used in this project, it is generally useful to pool the necessary areas of expertise among several individuals. It is helpful if one member of the team is familiar with survey research, application of survey sampling weights if applicable, and interpretation of test statistics.

An even simpler analysis of equity can be made with easily available data, either online or in printed reports. DHS data is available in interactive online tables through the StatCompiler service, and MICS data can be found through the UNICEF ChildInfo and DevInfo Web sites.[3]

Online tools alone can help users begin to design equity-sensitive health policy. For example, The Inter-American Development Bank (IDB) recently introduced an online MDG-based equity monitoring database that uses household survey data (although the number of health indicators included is fairly small).[4] These tools can help with singly-stratified analyses, and can be used to compare across countries and in some cases, across time. Similarly, DHS data may be mapped using freely available online GIS tools, such as STATMapper.[5]

Creating complex and simultaneously stratified tables requires some additional knowledge and access to statistical software such as SAS, SPSS, or Stata. Leveraging the GIS capacity of (freely available) proprietary software such as ArcMap or ArcView, or ArcExplorer, CSPro, or GRASS, will facilitate more powerful mapping.

---

[3] DHS: www.measuredhs.com; MICS: www.childinfo.org, www.devinfo.org. Special permission may be needed for access to some datasets.

[4] http://www.iadb.org/sds/xindicators/

[5] STATmapper is an interactive tool that produces maps of selected data from DHS surveys completed after 1998. Maps can be created for individual countries, or for zones, or groups of countries. Indicators available for mapping include: characteristics of respondents and housing units, fertility rates, childhood mortality rates, and measures of anemia and malnutrition (www.measuredhs.com).

Chapter 2
The Data Sources: DHS and MICS

DHS and MICS were chosen as data sources because in many countries, they are the only reliable data source for certain health indicators. These data sets are two of the leading international household survey programs and they report data on health, employment, education, family planning, and knowledge of disease. For most countries, these data sources (international household survey programs that report data on health, employment, education, family planning, and knowledge of disease) are readily obtainable from the Internet. A typical survey consists of a household questionnaire, with many additional questions about all resident women aged 15–49, children under five (addressed to the mother or primary caretaker of the child) and, for some DHS, a subsample of men age 15–49.

These surveys usually have large enough sampling frames to conduct simultaneously stratified analyses (see Table 1). DHS surveys usually include between 5,000 and 30,000 households while MICS are typically smaller, with 1,000–7,000 households. The sampling frame of these surveys often omits certain vulnerable populations, including orphans. DHS, and to a lesser extent MICS, provides a number of online manuals and tools to aid in interpretation and analysis.[6]

Geographic data

The critical geographic data required to map the health data presented here are the administrative boundaries associated with survey strata. These permit visualization of the spatial distribution of single indicators and the difference in distribution between different indicators or different populations within regions. Boundary data are generally not distributed with surveys, but they are increasingly available through independent and publicly available sources[7]. Increasingly in the last decade, surveys collect and distribute more precisely geocoded information about every survey cluster, or in rare circumstances, every household.

One of the most powerful features of a geographic information system is its ability to relate data sets on the basis of common geographic attributes. For example, climatic zones, population density, and relationships to roads can be assigned to each survey cluster, and therefore every household in it. These geographically derived attributes can then be treated like any other stratifier. For example, prevalence of immunization(s) can be stratified by classes of population density or proximity to road networks. In some countries, the locations of all public (and in less common cases, nearly all private) health facilities have been recorded in surveys or routine health management information systems, so proximity to health facilities can be measured. In all cases, data integration should be made with care, and considerable attention should be made to the spatial scales of the variables being combined so that the resulting data are meaningful for use as stratifiers.

---

[6] http://www.measuredhs.com/accesssurveys/tools.cfm

[7] These include the DHS STATMapper (see note 5) and global vector data of hunger and infant mortality and small area poverty maps from the CIESIN poverty mapping website, www.ciesin.columbia.edu/povmap. Related boundary data that may be useful in some circumstances are distributed by the World Health Organization's Second Administrative Level Boundaries project (SALB; http://www3.who.int/whosis/gis/salb/salb_home.htm), and several Consultative Group on International Agricultural Research (CGIAR) centers including the International Water Management Institute (IWMI; www.iwmi.cgiar.org), the International Potato Center (CIP; www.cipotato.org), and the International Center for Tropical Agriculture (CIAT; www.ciat.cgiar.org).

Other instruments

The Living Standards Measurement Survey (LSMS) is another survey amenable to health equity analysis.[8] Also household-based, it has a greater focus on price and income data. Data access ranges from easy to completely restricted, depending on the country. Other survey programs that may be of interest are the World Health Surveys (WHS), the Urban Inequities Surveys (UIS), the Center for Disease Control (CDC) International Reproductive Health Surveys (IRHS), and the Pan Arab Project for Family Health (PAPFAM).

Of course, many national survey programs and individual surveys provide valuable data that may be more appropriate for the circumstances and needs of a given community. For example, HIV/AIDS prevalence cannot be ascertained from most DHS and MICS surveys and yet is an essential health indicator to monitor in equity terms. This health data would need to be gathered from an HIV/AIDS survey or other reproductive health survey. Indonesia's Family Life Survey (IFLS) provides a wealth of information on health and other social indicators that is explicitly tailored to national circumstances. Lastly, a variety of development organizations carry out surveys within project sites, as broad or as narrow in scope as the projects that they document.

Table 1. Two Major Population-based Surveys: DHS and MICS[9]

	DHS Survey	MICS Survey
Commissioning Agency	USAID	UNICEF
Key Features	• Household questionnaire and women's questionnaire • Focus on nutritional status and anemia; mother's health; fertility; children's health; and family planning • Some DHS include service-availability questionnaires (infrastructure) • Some include biomarker information (HIV, syphilis, etc.)	• More child labor and same education information • Several modules which can be omitted or included (gives countries flexibility) • Includes many very poor countries not in DHS • Based on human rights principles. Includes same specific attention to malnutrition and mortality disaggregated by age, etc.
Number of Countries	70	65 (MICS2)
Sample Sizes	5000–30,000 households, nationally representative	1000-7000 households, nationally representative (some are larger)
Frequency	Typically every 5 years	Every 5 years

---

8 The LSMS is available from the World Bank online at: http://www.worldbank.org/lsms/guide/select.html.

9 Adapted from Komarecki (2003).

Strengths	• Geocoding available for many current surveys • Qualitative data, biomarker data, service provider assessments	• Lower cost, faster turnaround time • MICS3 is specific to MDGs (21 of the 48 MDG indicators) • UNICEF provides less direct technical assistance; carries out capacity building and training
Weaknesses	High cost ($0.5M) and long turnaround time	No geocoding yet. Information on HIV/AIDS not as complete as in DHS. Information on ethnicity to be included next round

In some data-poor countries, demographic surveillance systems (DSS) collect information on a continuous basis on health and its determinants, serving as limited vital registration systems for field research stations. Unlike standard vital registration data, DSS data are collected in order to evaluate health interventions aimed at reducing socio-economic differentials in mortality and morbidity (Ngom et al. 2003). Lastly, facility-based surveys, reviews of national health accounts, and health workforce assessments can provide important information about health systems and how resources allocation within those systems.

Chapter 3
How to Define Health Equity and Inequity

Health equity is based on simple notions of fairness and distributive justice. The nuances of this field and the distinction between the definition of disparity and inequity are briefly discussed here.

Certainly, some individuals experience challenges to survival while others do not. Unfortunately, some children will have stunted growth, or even die, while other children will grow up without any serious health issues. Some of these differences between individuals are unavoidable—the result of random variation. These may be termed disparities or inequalities, but may not necessarily be *inequities*.

However, when disparities are strongly and systematically associated with certain social group characteristics such as level of wealth or education, whether one lives in a city or rural area,[10] they are termed inequities. Put another way, disparities or inequalities are differences in health status between groups *independent of an assessment of fairness.* Inequities, however, are a subset of disparities *judged to be avoidable and unfair;* and thus require some sort of ameliorating policy intervention (Evans, Whitehead et al. 2001). *Health inequity* implies a failure to avoid or overcome inequalities in health that infringe on human rights or are otherwise deemed unfair. The concept of health inequity is a moral category deeply embedded in political reality and the negotiation of social power relations (WHO 2005). Conversely, health equity may be defined as:

> *the absence of unfair and avoidable or remediable differences in health among populations or groups defined socially, economically, demographically or geographically (WHO 2005)*

Some of the MDGs and related health targets are outcomes (e.g., child mortality) while others are inputs/processes (e.g., access to health care). For both types, empirical data reveals avoidable disparities between population groups—or inequities. For example, in Tajikistan there is a clear educational gradient in an access indicator, the use of contraception. Sixteen percent of women with no education, 26 percent of those with secondary education, and 41 percent of those with tertiary education use modern methods. Because this disparity is *avoidable and unfair*, it is identified as an inequity.

Arbitrary thresholds must sometimes be set to quantify disparities between groups—but it is always important to test for statistical significance. For instance, average weight at birth could be 2.4 kg in one province and 2.6 kg in another. The first figure meets the World Health Organization (WHO) definition of low birth weight (less than 2.5 kg) while the second figure falls above the official cutoff. However, these estimates are based on survey samples and reflect a range within each province. Statistical tests must be employed to establish which numerical differences actually represent inequalities.[11] It may be that neither figure is statistically significantly different from 2.5 kg, and thus, what appears to be an inequity is not.

This *Guide* recommends tests of statistical significance in order to allow definitive judgments about which differences between indicators are actually inequities. We note that the appropriate level of significance for testing, usually at a significance level of five or 10 percent, retains a certain level of arbitrariness

[10] We call these characteristics stratifiers, following standard terminology.
[11] See McCloskey and Zilliak (1996) for the need to complement statistical tests with common sense when evaluating what magnitudes are real differences or not.

Chapter 4
What to measure: *Health Indicator Selection*

The choice of indicator is critical because the degree of inequality revealed can vary greatly depending upon the indicator chosen (see Box 1). Not only do child health indicators differ from reproductive health indicators in degree of disparity, but disparity varies depending upon whether the indicator measures health status, access to health care, and care-seeking, knowledge, or opportunities to be healthy. Thus, careful selection of multiple indicators will yield a more complete picture of health inequity. Important general considerations include cost of data collection, data quality issues (such as the reliability of self-reporting), availability for monitoring at appropriate time intervals, cultural appropriateness, susceptibility to policy intervention, and level of technical capacity required for analysis. As mentioned above, equity analysis must be chosen based upon the epidemiological profile of the population.

BOX 1. Criteria for Selecting Indicators to Monitor Equity

The indicators must meet standard scientific and ethical criteria (validity, reliability, ethical and cultural acceptability), with data sources of acceptable quality and reviewed by local experts who know from experience the source's limitations.

Differences in the indicators between better- and worse-off groups should be relatively likely to reflect avoidable, unfair gaps in important conditions that could be narrowed through policy changes in any sector that influences health, not just health care.

Appropriately disaggregated data (disaggregated according to the social groups one wants to compare) on the indicators should be accessible for monitoring over time at the desired geographic level.

The indicators must occur frequently enough in the groups to be monitored to permit reliable estimate of differences between the groups.

The complexity of analyzing, presenting, and interpreting information on the indicators should be considered in light of the experience of the local analysts, decision makers, and other key participants who must perform these functions on a routine basis, generally without outside technical support.

A range of indicators is needed to reflect important aspects of health and its major determinants, including indicators of health status, major determinants of health, and health care itself (including financing, allocation of health care resources, utilization of health services and quality of health services).

When a large number of indicators meet the above criteria, an additional criterion to set priorities among the options to be presented to non-technical audiences should be the extent to which the indicators are likely to be meaningful (i.e., to reflect important concerns) to the public and policy-makers.

Source: Braveman (1998).

As the study of health equity and the social determinants of health develops, greater understanding of the particular health outcomes (and social stratifiers) most responsible for health inequities will become clearer (WHO 2005). This knowledge should inform the choice of indicator and analysis undertaken in efforts such as the one described here.

For the examples used in this *Guide*, indicators were chosen to align as closely as possible with the indicators chosen to track the child and maternal health MDGs (see Appendix A). They include outcomes (underweight, child mortality); access to care or preventative interventions (skilled attendant at birth, measles and DPT vaccination, contraceptive prevalence rate); knowledge (AIDS); and fertility-related women's status indicators (age at first marriage) (Wirth et al 2006).

Chapter 5
Across Which Population Groups?
Social Stratifier Selection

An equity analysis requires dividing a population into groups according to social advantage. The social stratifier most frequently associated with inequities is wealth or income. In many analyses, a score based on the presence of common physical household assets is used as a proxy for wealth, rather than for monetary income or expenditure. These scores are typically used to divide the population into quintiles. However, stratification by these wealth quintiles alone is not always the most appropriate way to measure inequities in health. In countries with extreme poverty, the wealthiest quintile often resides only within the capital city, so studies based on wealth alone may mask differences between other areas. Furthermore, household-level measures, such as family assets, ignore intra-household inequalities, such as those conferred by gender, age, or position within the household family structure. When appropriate data are available, these inequalities should also be evaluated.

Table 2 defines the following six stratifiers: wealth, sex, educational attainment, ethnicity (urban versus rural), residence, and region. Other potential stratifiers include occupation, employment status, age, birth order, and geographic regions defined by factors such as climate or remoteness.

Some of these stratifiers lead to a natural hierarchy, while others do not. For instance, within most countries the richest households would be expected to have better access to health care and health outcomes than the poorest households. Similarly, one would expect people with more education to have greater access to information and means of production than those with less education, and for this to be borne out in their health. However, without *a priori* knowledge of a country, it would be difficult to know whether a given region is better off than another. While some regions are better off than others in any country, and this may translate into differential health outcomes, a large region is likely to include people with a broad range of resources, from elites to landless poor. In some contexts discrimination by ethnicity is systematic and related to a long-term overclass/underclass structure, while in others it is not. Here, wealth and education are referred to as vertical stratifiers, and region and ethnicity as horizontal ones.

Wealth quintiles

As described above, wealth is usually measured based upon a set of household assets rather than in terms of monetary income or expenditure. The method for calculating a wealth index developed by Filmer and Pritchett (2001) has become standard. The equal fifths of the population (quintiles) are ranked from 1 = poorest to 5 = richest.

Table 2. Definitions of Stratifiers and Adjustments Made

STRATIFIER	DEFINITION	ADJUSTMENTS
Sex	Sex of child	
Educational Attainment	Mother's highest level of education	Grouped into None, Primary, and Secondary. Non-formal curricula and strictly religious education excluded
Residence	Urban or Rural	
Ethnicity	Country-specific	Uses standard DHS recodes (not available in MICS)
Ethnicity (recoded by Group Dominance)	Country-specific	Divided into dominant, non-dominant, and secondary dominant (where available based upon country-specific literature)
Wealth	Quintiles of wealth (country-specific)	Ranges from 1 = poorest to 5 = richest
Poverty line	Above or below national poverty line	Data on percent of the population below the poverty line (*Human Development Reports*) applied to wealth index data, to group households into 'poor' (below poverty line) and 'not poor' (above poverty line)
Region	Country-specific	

Wealth by poverty line

A variant of wealth quintiles, the 'Wealth by poverty line' variable can be created using an existing wealth index. This approach was developed as a very simple policy-relevant distinction between not-poor and poor (Sahn and Stifel 2000). Using external data on the percentage of the population living below the poverty line, one can rank surveyed households by the wealth index, defining this percentage at the bottom of this ranking as either poor (below the poverty line), and the rest as not poor (above the poverty line). Data from other surveys or studies on the percentage of the population living below the poverty line can be used to classify households as 'poor' based on the wealth ranking. If *x* percent of the population lives below the poverty line, the bottom *x* percent of the households ranked by wealth can be considered 'poor', and the rest as 'not poor'.

Sex

The sex of a child is its biological assignment as male or female. This is used as distinct from ‘gender,’ which encompasses socially constructed differences between the sexes. Because gender roles vary across societies, the term ‘sex’ is used in this document. However, health inequities can be due to either. For example, there is no known medical reason girls should be immunized less than boys. However, there may be certain biological differences in growth between girls and boys that could explain some of the differences in underweight status.

Educational attainment

Educational attainment is a factor in the health of both adults and their children. Here we are concerned only with the highest level of education attained by women, broadly categorized as None, Primary, and Secondary or more. It is an imperfect proxy of a woman’s empowerment, which is in turn linked to her ability to promote her welfare and that of her children. Other aspects of gender inequality, such as measures of household decision-making or mobility may affect women’s ability to receive health and family services, and where data permit, should be included in an equity analysis.

Ethnicity

In many countries, ethnicity is a profoundly inflammatory subject, and it may not be politically feasible to officially track health outcomes by ethnic group. Race, language, geographic location, economic class, or other attributes that align with ethnic divides in certain situations can be used as proxies to help collect data or design targeted programming. However, language and other proxies can also mask the diversity within an ethnicity. Where possible, ethnic data should be collected directly, and the heterogeneity of any given group or community should be taken into account.

Here categories provided by DHS are used to examine stratification by ethnic group, where data was available. In addition, the population is divided into three types of ethnic groups, which are referred to as ‘ethnicity recodes’: dominant, secondary dominant, and non-dominant. This exercise was based upon a brief and selective review of literature on the subject and was done for the purposes of illustration only.[12] Within a given country, better qualitative work and local expertise should guide the classification of ethnic groups into larger aggregate categories if necessary or useful for analytical clarity or large enough sample sizes.

Geography (Residence and Region)

The regions used here are survey strata[13] —the smallest geographic areas for which estimates are produced. The number of regions per survey varies with the size of the sample, the size and administrative division of the country or area, and other factors. Because there are typically more regions than sexes or education levels, sample sizes in individual regions can become too small to yield meaningful results when used in combination with another stratifier.

---

[12] These include Moyo (2004) for Ethiopia, and Brockerhoff and Hewitt (2000) for Ghana and Kenya. For Cambodia, a rough proxy was created as follows: residents of the two provinces with ethnic minorities in the majority, as shown in Elledge et al. (2001), were designated ‘Not dominant’. Both provinces are more than two-thirds minority, while all other provinces of Cambodia are less than one-tenth ethnic minority. No ethnic group information was available for the Dominican Republic or Tajikistan.

[13] In some cases, the actual strata are these units further subdivided by urban/rural, but this division does not result in contiguous regions.

Disaggregating data by region or by urban versus rural residence can be valuable, especially in combination with other stratifiers. However, disaggregation at this level is too coarse to highlight the disadvantages suffered by people living in peri-urban areas and urban slums. Some of the most profoundly dire circumstances coexist side by side with great advantage in cities. In such cases, rapid urban assessments may provide more information than census-based household surveys.

In the future, information about the location of each survey cluster will become available digitally. Further questions about geographic differentials may be asked with greater flexibility, at that time, but at present, this type of analysis is beyond the scope of this *Guide*.

Occupation

Occupation is an important stratifier of many of the indicators examined but was not included in the examples in this *Guide*. The categories used by DHS and MICS are not particularly useful for equity analysis; professional titles or types such as 'sales' could refer to an informal street vendor or a manager of a large department store. Clarification of the type of work and conditions associated with it would be more useful, although this may pose sensitive reliability issues. Items of greater relevance include: the continuity of work (seasonal, permanent, temporary); whether the job is in the formal or informal sector; the level or type of skill involved; and in the agricultural sector, land tenure.

Age

Age is an important factor for some indicators. The specific vulnerabilities and needs of adolescents should receive special consideration, particularly in the context of HIV/AIDS and the large relative size of their cohort in the developing world. For example, analysis of contraceptive prevalence rates (CPR) may benefit from disaggregation by age to examine the patterns of adolescent and others' access to contraception, particularly because of the greater chances of maternal mortality for adolescents.

Birth order

Birth order is an important stratifier for certain child health outcomes in some cultures. An analysis of health based on birth order can provide insight into relative advantages or vulnerabilities of being born first, last or somewhere in between. Needs and customs may require the eldest child to work rather than attend school, perhaps making them more likely to be disadvantaged. Conversely, eldest children may receive a larger share of the resources or inheritance than younger siblings, putting them at an advantage.

Groups that may be overlooked

As mentioned above, certain population groups may be completely overlooked in analysis that relies on population-based surveys or even vital registration. For example, these data sources rarely capture refugee populations—yet in many countries they represent sizable populations with poor health outcomes and inadequate access to health care. Orphans, foster children, street children; and hospitalized, institutionalized, and incarcerated individuals are also outside of the scope of many surveys. Other sampling issues include the undersampling of urban areas (especially slums) in the presence of rapid urbanization, language barriers between enumerators and respondents, and underreporting of certain conditions due to stigma.

PART II
Chapter 6
Constructing Singly-stratified Tables

Simple explanation

Singly-stratified tables are simple cross-tabulations that show how a health indicator varies across different social strata. For example, such tables may reveal how measles immunization varies across regions, levels of maternal education or between ethnic groups. The tables constructed for this analysis used 20 indicators and six different social stratifiers.

Technical explanation

DHS and MICS data were downloaded[14] in SPSS format, and relevant variables were 'cleaned.' Variables were recoded to exclude missing or irrelevant data, and some answer choices were regrouped to ensure large enough samples or to standardize across countries. Standard DHS recodes and indicator definitions[15] are often appropriate, but must sometimes be adjusted in order to accommodate the conditions of a given country or region.

To calculate each indicator, cases were divided into each category. Where possible, calculations were verified against tables in published DHS and MICS reports. Figures based on small sample sizes (250 births for mortality calculations, 25 cases for all others) were suppressed, and those based on relatively small sample sizes (500 births for mortality calculations, 50 cases for all others) were marked[16]. All data were copied into Excel, creating pivotable tables so that dynamic comparisons by indicator, stratifier, and country could be made (see Appendix B for more detail on pivot tables). To facilitate printing and distribution, a static version was also created for each country. Table 4, as an example, is the static table showing child health indicators for Kenya.

Each singly-stratified table presents the indicators described at the top for each stratifier at the left. In nearly all cases, the numbers represent percentages of the class described at left that fulfill the requirement shown at the top (e. g., 'having a DPT2 vaccine,' 'being underweight,' or 'using a modern form of contraception'). Exceptions to this are the columns 'Age at first intercourse' and 'Age at first marriage,' which represent mean values for the group at left.

Calculation of mortality rates is also somewhat more complicated. (See Appendix C for a brief discussion of issues concerning the calculation of child mortality.) For each stratification class (e. g. sex, region), a difference test is also reported (labeled *p*-value in gray italics in the first line of the relevant cell). These are probabilities of the null hypothesis that the values of an indicator for each level within a stratifier—'none,' 'primary,' and 'secondary' within education—are not significantly different from each other. Tests of significance were not performed on the mortality rate indicators, because they are rates rather than proportions. However, standard errors (SE) are given in individual DHS reports for the national mortality rates. National-level SE give some indication of which of the differences between groups of similar size are likely to be considered significant.

Singly-stratified tables were compiled from DHS surveys of Cambodia (2000), Dominican Republic (2002), Ethiopia (2000), Ghana (1998), and Kenya (1998), and a MICS survey of

---

[14] Data for most DHS surveys are available upon request, with a very brief description of research objectives, at www.measuredhs.com. MICS surveys may be found at www.childinfo.org.
[15] See Rutstein and Rojas 2003
[16] These thresholds follow DHS conventions for reporting small sample size.

Tajikistan (2000). These countries and surveys were chosen because they are selected Millennium Project case studies. Tajikistan data were taken directly from aggregate tables distributed by UNICEF, whereas most measures from DHS are calculated from individual woman- or child-level data. The remaining data were copied from DHS reports or the DHS website. When possible, indicator definitions were harmonized across the five DHS countries.

Some indicators reported here differ from those in official DHS reports (Rutstein and Rojas 2003). For example, values of ‘don’t know’ or ‘missing’ were excluded from the examples used here, whereas in DHS reports, these categories are sometimes explicitly reported, or, in some types of AIDS knowledge, considered equivalent to ‘no.’ Similarly, the DHS reporting of contraceptive prevalence is based on rates for women currently in union, whereas the examples used here are based upon report for all women. Finally, the examples used here report mean age at marriage, as opposed to median.

Tests of significance

In working with differences in sample means—i.e., the values obtained from the survey data–it is important to correctly identify statistically significant differences between groups in the population as a whole. In other words, we want to know whether the difference between reported values for two groups reflect the existence of two truly different values or simply are evidence of random variation within one large group that happens to make the value for one subgroup slightly larger than that for the other. To answer this question, tests of significance must be conducted.

Two different kinds of tests are considered. Direct comparisons determine whether two groups have different values of a given measure. A direct comparison might test whether boys have different nutritional status from girls. Omnibus tests determine whether a population grouped by a stratifier with more than two groups has, in general, different values between the groups. For example, an omnibus test of access to health facilities in Cambodia by region would test whether, in general, the 17 survey regions have different levels of access. Several different significance tests can be used, but this *Guide* uses Pearson’s *chi*-squared, which is widely available in statistical software packages as an option in cross-tabulation functions. The resulting test statistic is converted to an F-statistic accounting for survey design (here, the sample weights) that yields the relevant significance value. This can all be carried out automatically in one step in most statistical packages.

In the direct comparison, like with other significance tests, the output is a probability, or level of confidence. In this case the output is the probability that the value of the indicator is the same for two groups in the population as a whole. Thus, the closer the probability is to zero, it is less likely that the value is the same for the two groups (i.e. it is more likely that the difference is significant). Note that the significance is not dictated by the size of the difference between the group averages – it is also influenced by the variation within the groups and their relative sizes.

The absolute size of the disparity between groups is not necessarily an indication of its level of significance. In Cambodia, for example, the average age at first intercourse is 19.4 for the dominant ethnic group, and 18.6 for others, a seemingly minute difference. The difference is 0.8, or less than 5 percent of the mean, and yet it is significant at the 0.5 percent level. This is related to fact that the dominant group represents the overwhelming majority of the population, and because the within group variance in age at first marriage is smaller. Conversely, 23 percent of boys and 20 percent of girls receive DPT vaccinations in Ethiopia, but this approximately 15 percent difference is not significant even at a 10 percent level of confidence.

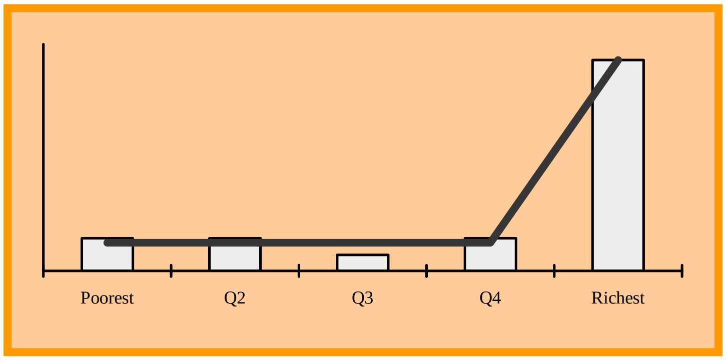
For an omnibus test to be significant, not every region needs to be unique. Two particular regions could have identical access values, but if the other 15 regions differed greatly, then region would be a significant stratifier. Similarly, if only one of the 17 regions is significantly different from all the others (yet all the others were alike), the omnibus test would also indicate there are significant differences amongst the population as a whole. In other words, the omnibus test highlights the presence of any significant disparity, but does not consider pair-wise (i.e., any two regions) tests or reveal much information on the nature of the distribution of values.

This implies that further probing might be needed. Once it is established that [not all] regions vary significantly in their levels of access, it might be important for policy purposes to find out if all regions are different, whether only one is completely different from the rest, or whether the heterogeneity is more complex.

Graphical visualization or 'eyeballing' the numbers in the table should be sufficient to establish a few simple hypotheses about the structure of differences. For instance, a whisker plot could be used, placing all the regions along the horizontal axis. As the whiskers could be made to represent the range and the standard deviation for each category (region in this case, but it could also be wealth quintiles, ethnic groups, etc), the overlap among the whiskers could give an indication of which regions to clump together and which ones to select for further testing.

Bar charts can also be used for this purpose. Figure 1 measures the indicator Contraceptive Prevalence Rate-Modern Methods (CPR-MM) from the poorest to the richest quintile. Although the first four quintiles are similar, the fifth and richest quintile stands out as dramatically higher. An omnibus test would indicate not all five quintiles are the same.

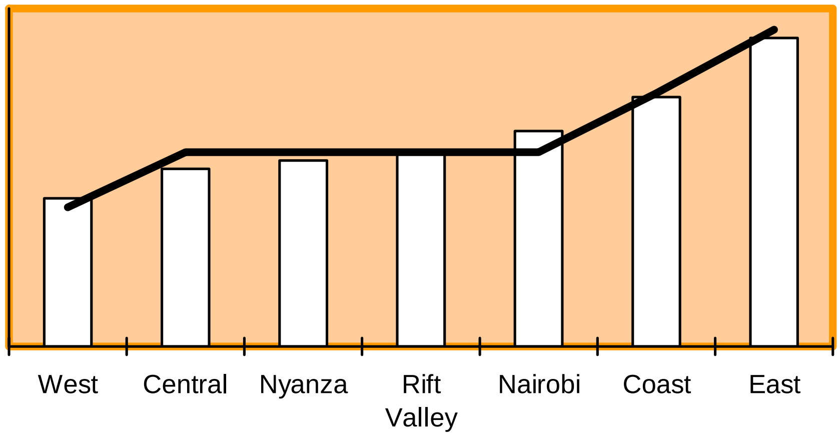
Figure 1. CPR-MM by Wealth Quintile, Ethiopia



A second test is needed to establish whether the fifth and the fourth quintile are different. If they are, and this can be established by a direct comparison of Q4 and Q5, an omnibus test can be used to test if the rest (Q1 through Q4) are all similar. If they are, the pattern represented by the thick line emerges.

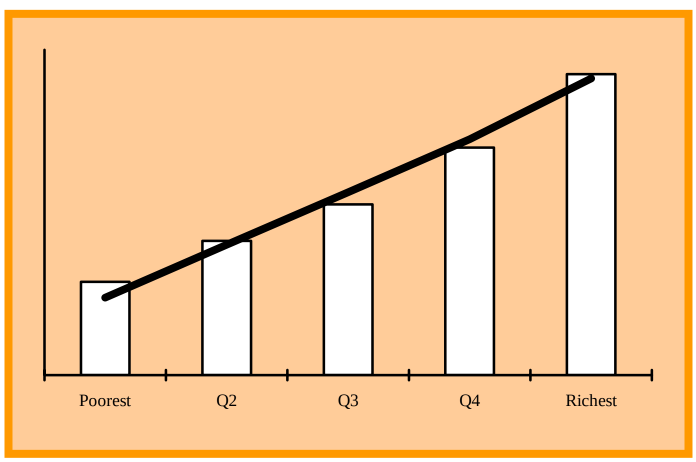
This pattern is different from the one depicted in Figure 2. In this case, if the omnibus test indicates there are differences among the regions (i.e. the test gives close to zero probability, meaning it is unlikely that the access to health services is the same across regions), a second omnibus test can be performed only for the Nairobi, Coast and East regions. This can be complemented with direct comparisons of the North and the South (or, if the omnibus test validates the similarity, with either Center or East) as well as West and South. The pattern, again, could be represented by the thick line.

Figure 2: Access to Health Facility by Region, Kenya



Another interesting case occurs when the omnibus test points to the groups being different and the graph suggests a gradient (see Figure 3). In this case, the second round of tests should include a series of direct comparisons (A and B, B and C, etc). However, this needs to be done carefully. It is possible that each direct comparison is not statistically significant (i.e. A is deemed not different from B, which is not different from C, etc) but certainly A and E are different. That is the reason why the omnibus test indicated some differences were significant. Then, further direct tests are needed (e.g., A and E, C and E, etc).

Figure 3. Skilled Birth Attendant by Wealth Quintile, Kenya



Tests of significance are an important tool for the identification of patterns and associated policy implications. The first example reveals that a single group difference drives the finding of disparity. The second example displays a more complex pattern with extremes in the North and West and similarity in the three middle groups. Finally, the third example reveals a linear trend in which the difference between each group is significant. The relationship between the stratifier and the indicator reveals the structure of the health inequities. Such information must inform the policies developed to reduce disparities.

Chapter 7
Reading and Visualizing Singly-stratified Tables

The singly-stratified tables described above are useful for highlighting the extent of health inequities between population groups. There are two main ways in which the tables can be used. First, the tables can serve as a reference to answer very specific questions about a particular indicator and stratifier. Second, the tables can be used to more extensively explore a question about the nature of disparities in a particular country, for a particular set (or all) indicators and for a particular set (or all) stratifiers—in order to efficiently answer such questions, the data may need to be repackaged into different visual formats. Each of these options is examined in this chapter.

Simple reading of singly-stratified tables

Specific questions about inequities between specific population groups *for a specific indicator* might include the following:

- *Do urban and rural children get vaccinated at the same levels in our country (or region)? What about male and female children?*
- *To what extent does maternal education make a difference in percentage of children who are severely underweight?*
- *Is there a gradient across wealth quintiles in the distribution of underweight children in our country (or region)?*

- ***Do urban and rural children get vaccinated at the same levels in our country (or region)? What about male and female children?***

Let's take Kenya as an example. A simple bar chart can summarize rural and urban outcomes for measles immunization (see Figure 4). This bar chart is based upon data from just one indicator and one stratifier in the singly-stratified tables. The chart highlights two points. First, the urban advantage in measles immunization is neatly depicted in this visual format. Second, it is clear that the population is disproportionately rural because the bar for the national average is closer to the rural bar. However, if only the bar graph is employed it is not possible to determine whether the urban/rural gap is statistically significant. Table 3 includes richer detail including three other immunization indicators and 2 other stratifiers as well as the *p*-values which assess level of statistical significance of the results. Table 3 makes it is clear that the *p*-value is 0.00 and thus verifies the significance of the gap. Beyond this quantitative interpretation, important policy questions—such as how to shift resources to rural immunizations—must be tackled.

To answer the rest of the first question, we find 'sex' in the first column. We then find that the first four columns are 'measles,' 'DPT1,' 'DPT2,' and 'DPT3.' The *p*-values are all well over 0.1, demonstrating that none of these differences are statistically significant at even the 10 percent level. Of course, if we were to look at sex differentials within specific regions or grouped with stratifiers, a more nuanced picture could emerge; this can be achieved through simultaneous stratification which is addressed in Chapters 8 and 9.

Figure 4. Measles Vaccination by Location, Kenya

Table 3. Immunization by Sex, Region and Residence, Kenya

STRATIFIER		INDICATOR VALUES			
Class	**Level**	DPT1	DPT2	DPT3	MEASLES
Sex	**P-value*	*0.98*	*0.51*	*0.32*	*0.77*
	Female	96	90	78	79
	Male	96	91	81	80
Region	**P-value*	*0.12*	*0.00*	*0.00*	*0.00*
	Central	100	99	98	92
	Coast	97	91	82	92
	Eastern	97	94	87	87
	Nairobi	(95)	(91)	(75)	(95)
	Nyanza	92	81	66	62
	Rift Valley	96	92	85	85
	Western	94	89	72	66
Residence	**P-value*	*0.40*	*0.66*	*0.88*	*0.00*
	Rural	95	90	80	77
	Urban	97	91	79	94

Table 4: Single Stratification for Eight Health Indicators and Eight Stratifiers, Kenya[17]

		Indicator Values							
		DPT1	DPT2	DPT3	Measles	Moderately underweight	Severely underweight	Underweight	Deliv. Assist. by SBA
Stratification Class	**Stratification Level**	**96**	**90**	**80**	**80**	**16**	**5**	**21**	**45**
***Nat. Avg.**		*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Education	**P-value*	89	79	68	70	24	10	34	27
	None	96	90	79	77	18	6	24	36
	Primary	98	95	87	90	9	2	11	72
	Secondary or more	*0.04*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Ethnicity	**P-value*	97	93	83	83	20	4	25	35
	Kalenjin	97	90	80	86	19	6	25	39
	Kamba	100	100	94	98	8	2	10	71
	Kikuyu	94	84	79	84	20	5	24	45
	Kisii	92	86	70	65	17	4	21	36
	Luhya	92	82	63	55	14	8	22	37
	Luo					(8)	(10)	(17)	(25)
	Masai	99	98	93	91	13	6	19	72
	Meru/Embu	96	93	84	90	19	12	32	27
	Mijikenda/Swahili					21	1	21	46
	Other								
	Somali	(100)	(100)	(95)	(86)	21	3	24	41
	Taita/Taveta	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Eth. recode	**P-value*	100	100	94	98	8	2	10	71
	Dominant - Primary	97	93	83	83	20	4	25	35
	Dominant - Secondary	94	88	76	75	17	6	23	41
	Not dominant	*0.98*	*0.51*	*0.32*	*0.77*	*0.11*	*0.69*	*0.10*	
Sex	**P-value*	96	90	78	79	15	5	20	
	Female	96	91	81	80	17	6	23	
	Male	*0.12*	*0.00*	*0.00*	*0.00*	*0.07*	*0.00*	*0.00*	*0.00*
Region	**P-value*	100	99	98	92	10	2	12	69
	Central	97	91	82	92	17	9	27	36
	Coast	97	94	87	87	16	7	23	48
	Eastern	(95)	(91)	(75)	(95)	10	2	12	78
	Nairobi	92	81	66	62	18	8	25	38
	Nyanza	96	92	85	85	18	4	21	39
	Rift Valley	94	89	72	66	17	4	21	33
	Western	*0.40*	*0.66*	*0.88*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
Residence	**P-value*	95	90	80	77	18	6	24	39
	Rural	97	91	79	94	9	3	12	71
	Urban	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Wealth by Pov line	**P-value*	97	94	84	87	11	3	15	58
	Not Poor	93	85	73	70	22	8	31	26
	Poor	*0.10*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
Wealth by Quintile	**P-value*	93	83	68	64	24	9	33	23
	1st quintile (lowest)	95	88	78	80	18	7	25	33
	2nd quintile	97	93	86	86	14	4	18	42
	3rd quintile	97	95	84	85	12	4	16	56
	4th quintile	98	95	85	90	7	2	9	80
	5th quintile (highest)								
Source: Kenya DHS (1998)									

17 Available online at www.ciesin.columbia.edu/Health_Equity_Tables.pdf

Table 4 aggregates a large proportion of the data generated for Kenya in a single table. This form of data presentation may be helpful for researchers seeking to display a great deal of information on multiple indicators and stratification levels in a succinct manner. However, simpler depictions of subsets of this data, as shown in Table 3, may be more appropriate for policymakers and visual presentations.

To answer the second question, we again assess the statistical significance of the data. For Kenya, we look at the first column of Table 4 to find maternal education and then look across to the column entitled ‘‘severely underweight.’ Here we see a more dramatic result in the data—the *p*-value is 0.00 meaning that the differences in this indicator (severely underweight) across categories of maternal education are highly significant, at greater than the 0.5 percent level.[18] We can then interpret differences between the three levels of maternal education, noting that the gradient goes in the direction one might expect (i.e. those with less education have higher levels of severely underweight children): 10 percent of children of women with no education are severely underweight, compared to six percent among women with a primary education and two percent of women with secondary or more education. In other words, women with a secondary education are five times as likely to have a severely underweight child as compared to those with no education.

In order to tackle the third question above, we refer to the first column of Table 4 and locate wealth quintiles as the stratifier and the three columns having to do with underweight as the indicator. In order to keep things simple, we will simply look at the overall ‘underweight’ category, though more nuanced analyses may wish to analyze ‘moderate’ and ‘severe’ underweight separately. As a first step, we note that the calculated *p*-value is less than 0.005 meaning that this set of data output is highly statistically significant. We see that nine percent of children are underweight in the fifth, or highest, wealth quintile, 16 percent in the fourth, 18 percent in the third, 25 percent in the second and 33 percent in the poorest. Thus, we can answer the question posed above: yes, ‘underweight’ reveals the expected gradient across wealth quintiles in Kenya as a whole, with each wealth quintile, moving from highest to lowest, yielding a greater proportion of children underweight, where one-third of the poorest children are underweight.

More complex visualization of data from singly-stratified tables

A broader inquiry might require more detailed analysis or even ‘repackaging’ of the data into more visually useful formats. Let’s take the following queries as examples:

- *Is one ethnicity disadvantaged across all (or most) health indicators?*
- *How do different types of immunization compare across wealth quintiles?*
- *Which regions suffer the greatest inequities in health?*
- *Across which stratifiers are inequities the greatest in a country (or region)?*
- *Which indicators reveal the greatest disparities?*

Results can be presented in a variety of ways, ranging from comprehensive tables showing numbers (see Table 4) to compact visualizations like whisker plots or spider graphs that pack

[18] Because only two decimal places are reported, ‘0.00’ means ‘less than 0.005’.

a great deal of information into a smaller space. Single indicators or stratifiers can be extracted from large datasets and presented by themselves. Countries' statistics can be compared across multiple stratifiers for one indicator or, alternatively, several indicators can be compared across one or more stratifiers. Following, we respond to the five queries listed above.

Table 5. Ethiopian Health Indicators by Ethnic Group

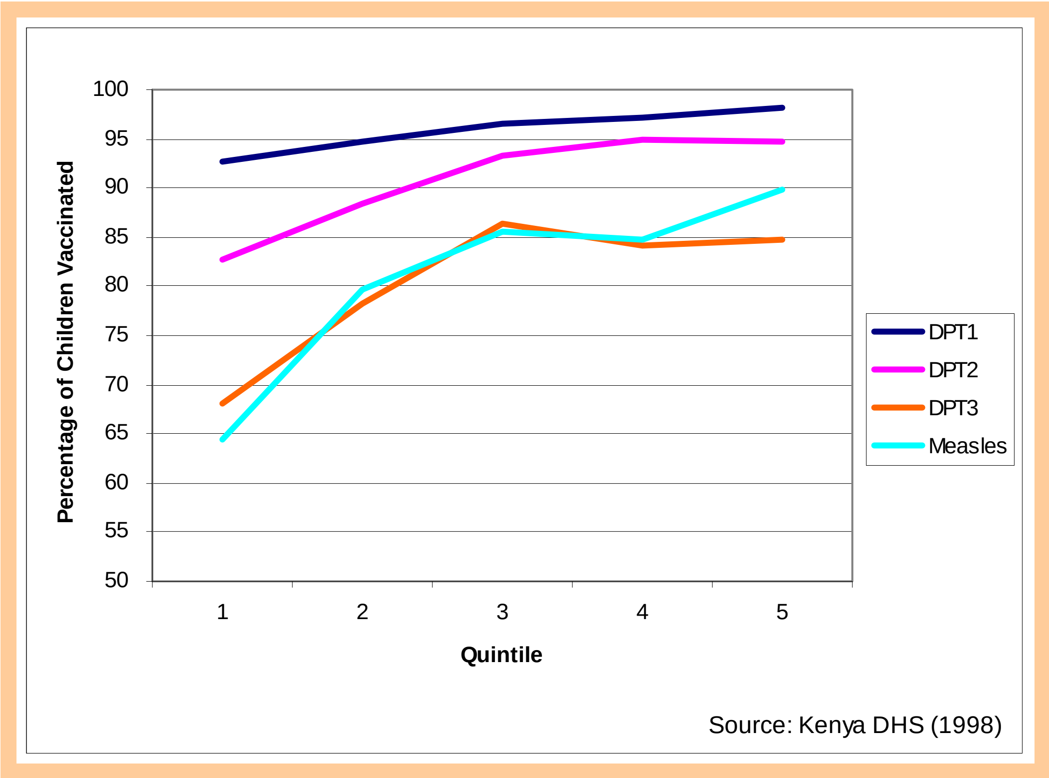
	% Access to a health facility (visited in last year)	Age at first marriage	% CPR - Modern Method	% Underweight	% Measles immunization
Ethnicity: ranked worst-off to best-off	Welaita (30)	Amharra (14)	Somalie (0)	Affar (55)	Affar (2)
	Affar	Tigray	Affar	Guragie	Welaita*
	Somalie	Affar	Welaita	Other	Oromo
	Other	Sidama	Other	Welaita	Other
	Amharra	Oromo	Oromo	Sidama	Guragie
	Oromo	Welaita	Sidama	Tigray	Amharra
	Guragie	Other	Guragie	Amharra	Sidama
	Tigray	Somalie	Tigray	Somalie	Somalie
	Sidama (49)	Guragie (18)	Amharra (8)	Oromo (43)	Tigray (67)
National Average	36	16	5	47	27

- ***Is one ethnicity disadvantaged across all (or most) health indicators?***

To answer this first type of question, we may want to construct a compact table for a few different indicators (see Table 5). Ethnicity significantly stratifies all the reproductive health and child health indicators (except 'moderately underweight'). The national average age at first marriage (16) obscures a wide range from 14 among certain ethnic groups (who are clustered in selected regions) and 18 among other ethnic groups. However, it is important to note that the relative advantages/disadvantages of the ethnic groups is not constant across indicators (see Table 5 for a subset of indicators). The Affar ethnic group is the exception—it is consistently well below the national average for all child and reproductive health indicators measured. And in general, the Tigray ethnic group is relatively advantaged, with the exception of AAFM.

By contrast, the Guragie group seems relatively advantaged for a few of the indicators (access to health facility, AAFM, CPR) but much more disadvantaged in child health (measles and underweight). Other groups like the Somalie, Welaita, Other, Sidama, Amharra and Oromo tend to vary in their level of disadvantage depending upon the health indicator. In short, the Affar seem to be the most consistently disadvantaged but that other groups are as or more disadvantaged for certain indicators.

Figure 5. Immunization by Quintile, Kenya



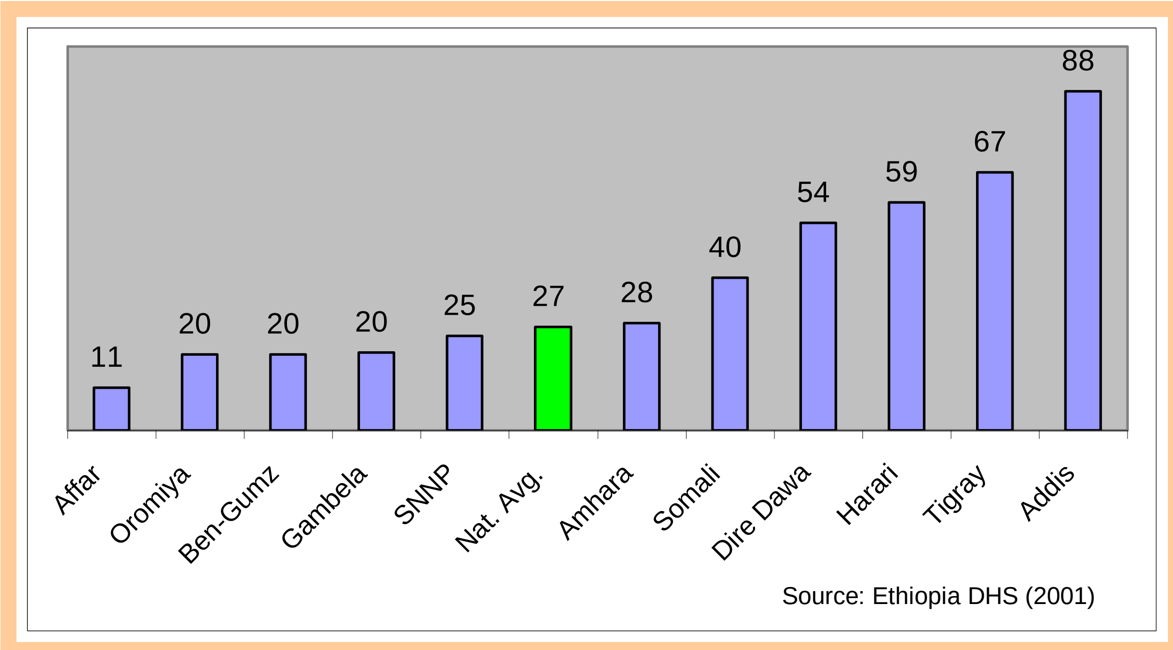
- ***How do different types of immunization compare across wealth quintiles?***

Clearly, different quintiles of the population have differential access to immunization. However, the 'shape' of these differentials varies depending upon the type of immunization. In Figure 5, DPT3 and measles have a much steeper gradient than do DPT1 or DPT2. And DPT3 and measles track each other very closely, evidence that may tell us something about the efficacy of the primary health care system (in comparison to mass immunization campaigns—or vertical programs). This evidence could be used for multiple policy discussions including a) the merits of mass immunization versus strength of the health system and b) how to reach the bottom two quintiles of the population.

- ***Which regions suffer the greatest inequities in child health?***

Tables hold answer to this question, but a bar chart such as the one displayed in Figure 6 depicts the range of coverage in a more striking fashion. Here we see measles coverage ranging from 11 to 88 percent across 12 regions in Ethiopia. Affar, Oromiya, Ben-Gumz and Gambela have extremely low measles coverage rates (20 percent or below). The next three highest regions still have very low measles coverage (25-28 percent) as compared to regions like Tigray and Addis. The national average is used here as a reference point, with about half the regions falling above this average and half below. A map would further illustrate these data, showing for example whether there are particular zones or corridors of regions that fall behind.

Figure 6. Children's Measles Coverage by Region in Ethiopia (percent)



- ***Across which stratifiers are inequities the greatest in a country (or region)?***

In answer to the fourth question, Figure 7 presents a whisker plot showing the proportion of underweight children by different social stratifiers. The effectiveness of this type of visualization, like others, depends in part on the number of classes defined by the stratifier. For example, dichotomous variables such as urban/rural and sex are less likely to have as wide a range of values as stratifiers with many categories like ethnicity and region (Anand et al 2001). Nevertheless, the whisker plot is useful because it succinctly captures both the level and the range of the indicator. In Figure 7, for example, we see that sex is not a very large stratifier of underweight in Ghana, but that education, ethnicity and region have a large portion of their 'range'/distribution above 25 percent underweight.

In a variation on the whisker plot visualization, Figure 8 also depicts a single indicator across different social stratifiers. This type of Figure includes more detail on the actual value of each level of stratification. So, for example, we can see, in a glance, that percentage of underweight children is stratified by about the same absolute amount for both dichotomous stratifiers (urban/rural and poor/not poor) but that the actual level of SBA for the poor is far lower than for rural dwellers.

Figure 7. Proportion of Underweight Children by Different Social Stratifiers, Ghana

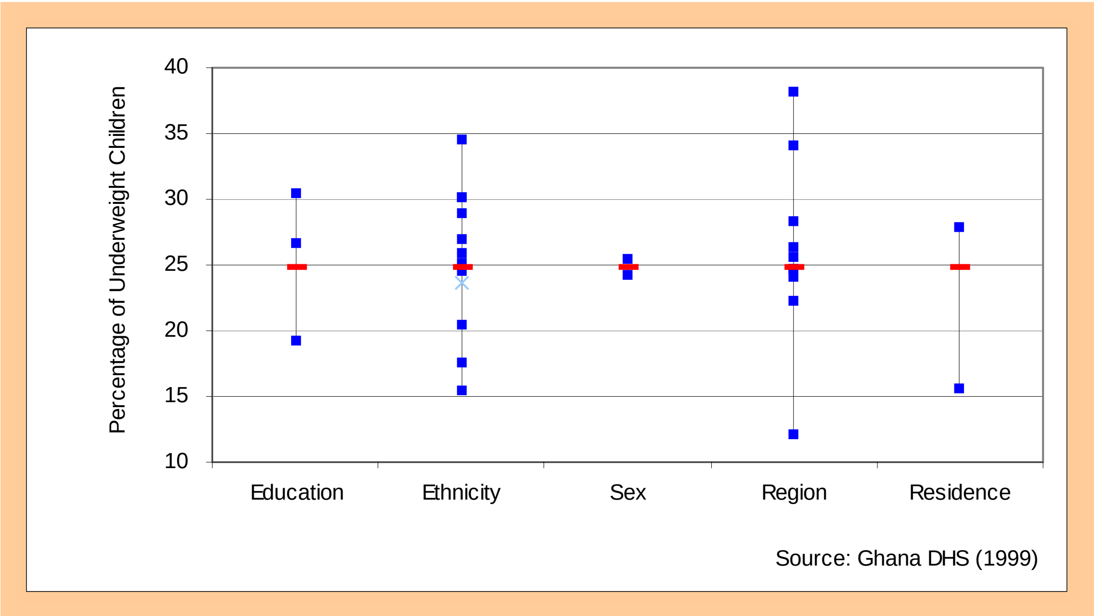
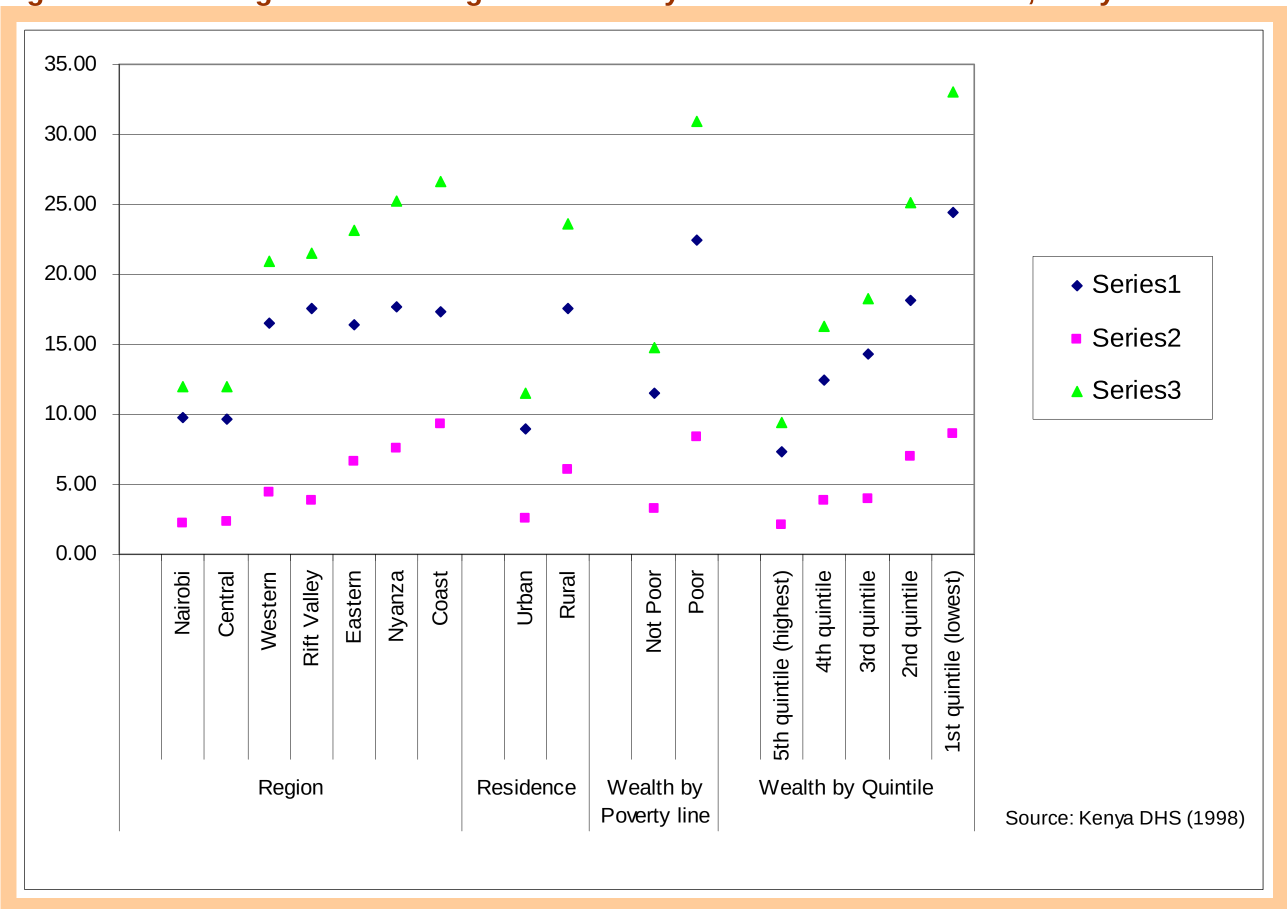


Figure 8. Percentage of Underweight Children by Four Different Stratifiers, Kenya



Note: Series 1 is moderate underweight, Series 2 is severely underweight and 3 is all underweight.

- ***Which indicators reveal the greatest disparities?***

In this final set of figures in this section, 'spider graphs' are used to display multiple indicators for a given country or multiple stratifiers for a given indicator. A spider graph can help to quickly visualize indicators which are more stratified than others as is shown in Figure 9. Here we see a spider graph of five child health indicators in Cambodia. Each colored line responds to a level of maternal education. A 'perfectly' distributed set of outcomes would be depicted with all three 'maternal education' lines fully overlapping. However, each level of education confers better health outcomes, particularly with respect to immunizations. Secondary education results in children's immunization levels higher than primary education, and in turn primary education results in levels higher than no education. As expected, the pattern reverses for underweight as secondary education results in the lowest levels of underweight—though the gap between primary education and no education is greater than the interval between secondary and primary. Furthermore, the spider graph reveals that disparities by educational level appear greater for the immunizations than for the underweight indicator.

In Figure 10, the same five children's health indicators are depicted on a spider graph, but the stratifiers are the child's sex, male and female. Here we see that the shapes of the 'male' and 'female' lines are much closer to coinciding. In fact, DPT 1 and DPT2 appear to be coinciding while for measles, DPT3 and underweight, females seem to be lagging a bit. In order to test the statistical significance of this pattern, the *p*-values from the tables would be needed.

Figure 9. Spider Graph of Child Health Indicators by Level of Maternal Education, Cambodia

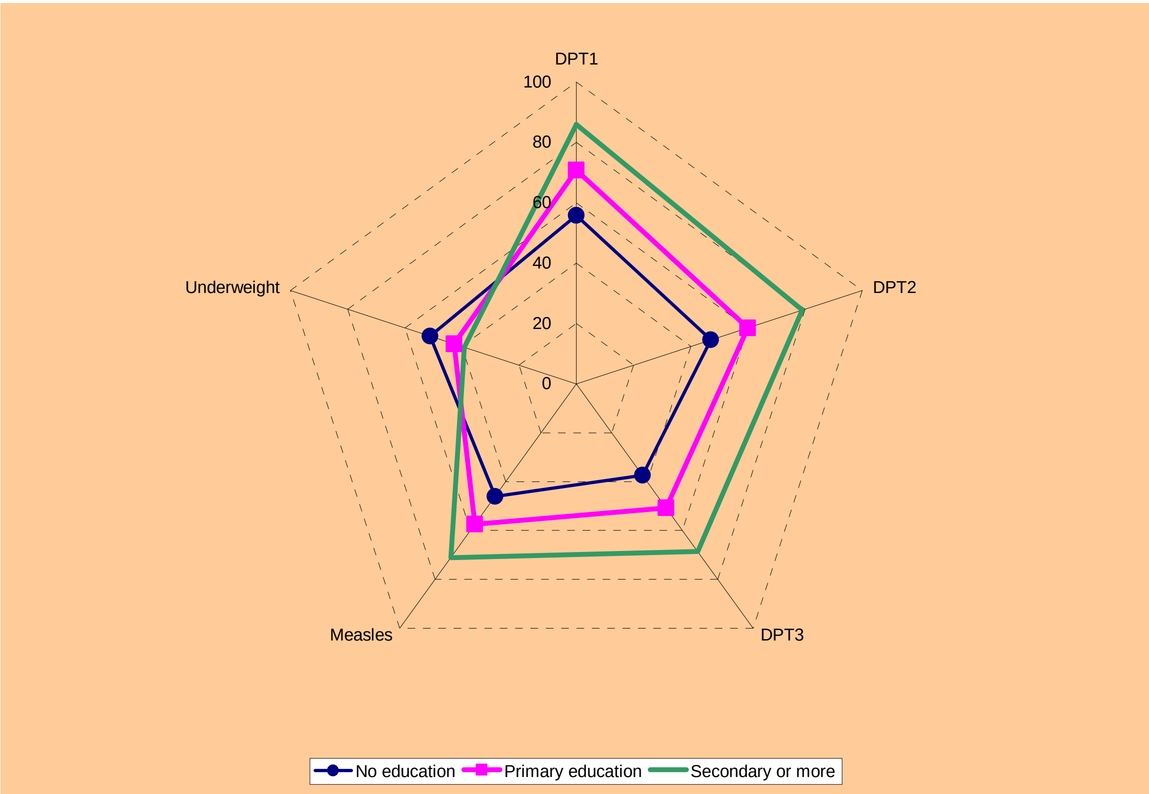


Figure 10. Spider Graph of Child Health Indicators by Sex, Cambodia

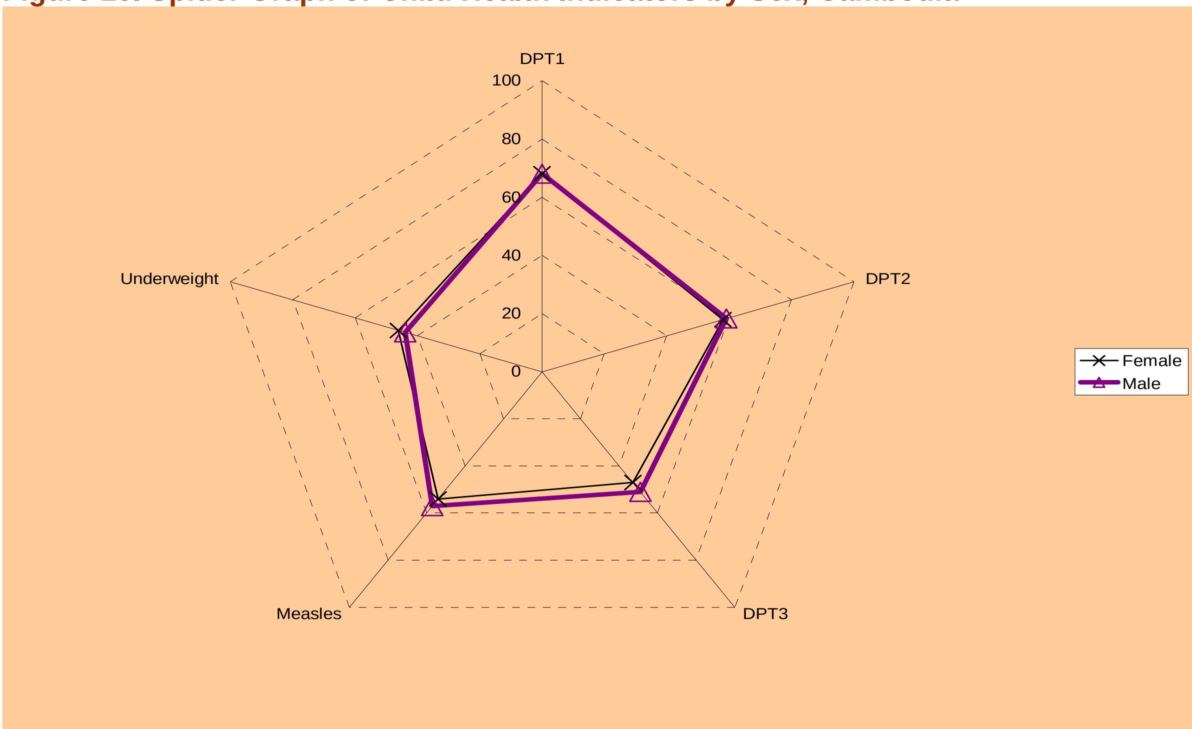


Figure 11. Spider Graph of Reproductive Health Indicators by Level of Maternal Education, Kenya

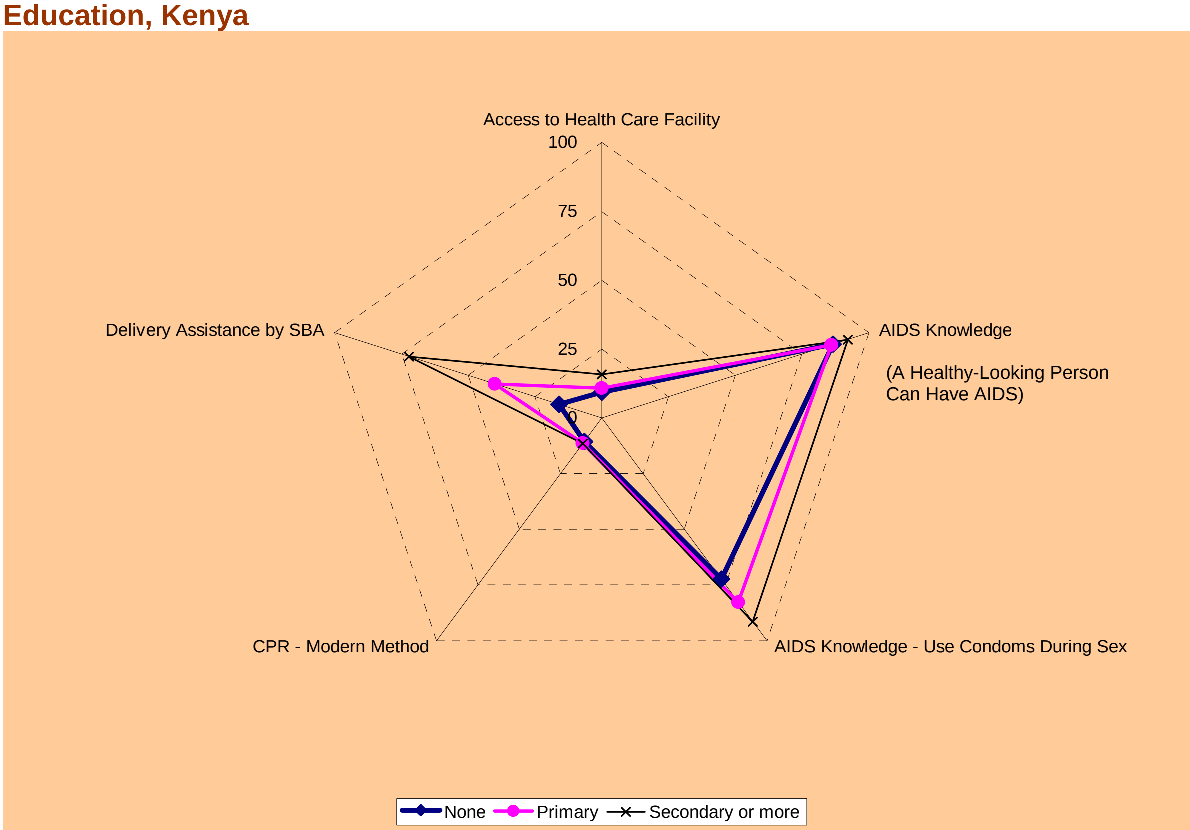
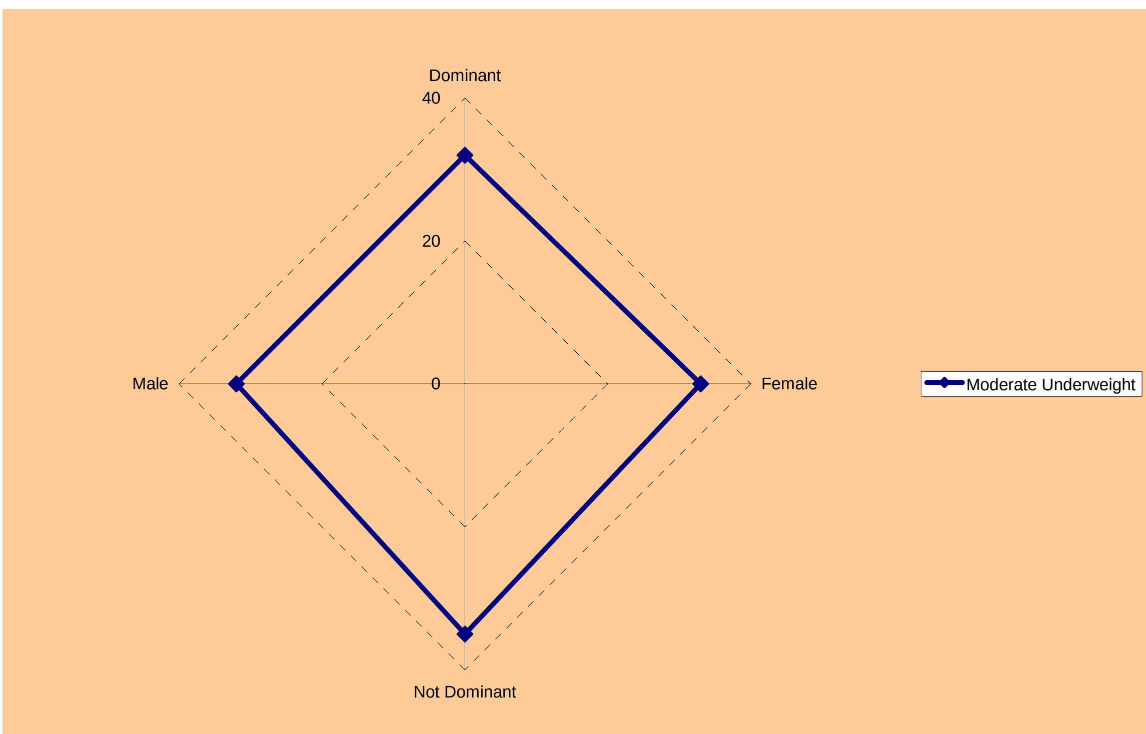


Figure 11 takes a set of reproductive health indicators from Kenya, using each 'spoke' of the spider graph as a different indicator and the 'strands' of the web to represent levels of maternal education. The indicators depicted here may be grouped into those which are 'knowledge' and those which measure 'access to care.' In the very skewed shape of this web, it is clear that relatively high levels of knowledge (across all educational levels) about AIDS is coupled with very poor and highly inequitable access to care (indicators such as access to a health facility and CPR-modern method). 'Skilled birth attendant' is the indicator with the most obvious and pronounced disparities by level of maternal education.

Another kind of story can be told using the spider graphs. Instead of using the spokes of the spider graph for different indicators, the spider graphs can be used to depict different stratifiers for a single indicator (see Figure 12). In Cambodia, for moderate underweight, there is no statistically significant difference between the dominant (32 percent) and not dominant (35 percent) ethnic groups. Similarly, there is no disparity between males (32 percent) and females (33 percent). However, as will be presented in the following chapters when simultaneous stratification is used, there appear to be significant differences between these groups.

Note the importance of the shape of a figure in which the spokes of the spider graph signify different stratifiers. The closer to an equitable distribution, the more symmetrical the shape will be. Here we see a near perfect square, indicating that the indicator 'moderate underweight' is distributed relatively equitably across male/female and dominant/not dominant ethnic group.

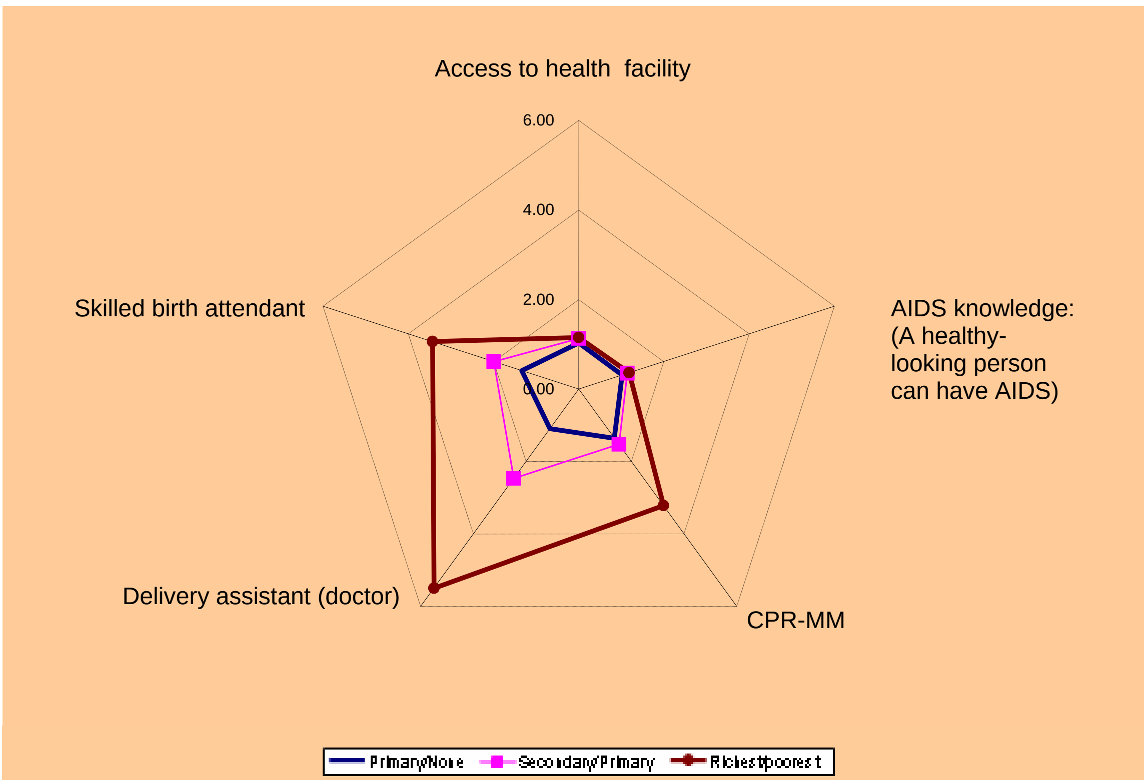
Figure 12. Spider Graph of Moderate Underweight by Two Stratifiers, Cambodia



Finally, as seen in Figure 13, rate ratios may be used to capture the degrees of magnitude of the gaps between different social groups. In this Figure, a set of Kenyan reproductive health indicators is depicted with rate ratios for wealth, residence and maternal education on the strands of the web.

The line which approaches 6.00 on the spoke for ‘SBA’ indicates that there is an almost six-fold greater access to a doctor as a skilled birth attendant for the least-poor quintile as compared to the poorest quintile of the population. This rate ratio decreases to about 3.5 for SBA (not doctor). This same figure also allows us to compare the rate ratios for secondary versus primary education as well as primary versus no education. The secondary/primary rate ratio is higher than the primary/none for the ‘SBA’ indicator, implying that the gap between primary and secondary educated women and access to SBA is wider than the gap for those with primary education as compared to those with none. By contrast, the rate ratios for maternal education groups for the AIDS knowledge indicators, CPR and access to a health facility all appear to be very close to one another. Note that this same technique of depicting rate ratios may be used for other stratifiers like residence, dominant/not dominant ethnic group, etc.

Figure 13. Spider Graph of Maternal and Reproductive Health: Rate Ratios for Two Stratifiers, Kenya



Chapter 8
Constructing simultaneously-stratified tables

Simple explanation

The simultaneously stratified tables build upon the simple stratification by adding a second social stratifier. Through this method the simultaneous effect two stratifiers on health indicators can be gauged. For example, the combined effect of sex and ethnicity on percentages of children immunized is revealed. The method allows comparison between health outcomes for girls of different ethnicities, or comparison boys and girls within the same ethnicity. Each pair of stratifiers is examined for each indicator (sample size permitting) so ethnic group is sub-stratified by sex, residence, maternal education, and wealth quintile; sex is sub-stratified by region, residence, maternal education and wealth quintile, and so on.

Technical explanation

Cleaning the data
The procedure for cleaning the data is similar to that used in simple stratification. However, because the number of categories expands geometrically with the number of dimensions of stratification considered at the same time, small sample size is more likely to be a problem. For example, the cross-tabulation of ethnicity and region is especially likely to result, in most cases, in many categories that are either empty or too small to produce valid estimates. However, ethnicities can sometimes be usefully recoded into dominant and not dominant classes with large enough samples for simultaneous stratification.

Tests of significance
The tests of significance are similar to those used in the singly stratified tables. However, because we are using two stratifiers, the tests can be run comparing rows or columns. For example, in Table 6, the upper leftmost test statistic (*p*-value is 0.0506) shows that, among mothers with no education, there is a significant difference between girls and boys' DPT3 vaccination. Analogous statistics comparing columns are calculated for each 'row'—i.e. for those with primary education and those with secondary education. As we can see, the null hypothesis that there is no difference in the immunizations rates of boys and girls *cannot* be rejected for mothers with primary education.

Another set of statistics is run within each column comparing rows. These answer the question of whether, among boys (or among girls), there is a significant difference in immunization by level of mother's education. For both sexes, the null hypothesis, that there are not significant differences between the specified groups is rejected at the 0.005 percent level, meaning that for both boys and girls there are highly significant differences by maternal education.

Gathering results
The method of simultaneous stratification (trivariate analysis) presented here was used to run each pair of stratifiers using each health indicator for each of six countries. This process resulted in the creation of a dataset with over 14,000 valid cells.[19] For example, in Kenya, ethnic group values are sub-stratified by gender, residence, wealth, etc., and so on for each combination of stratifiers, to determine the compounded effect of dual forms of vulnerability. As noted above, some pairings

were not generated in the simultaneously stratified analysis because sample sizes would result in the majority of classes being null.

Table 6. Immunization in Ethiopia Stratified by Maternal Education and Sex

	DPT 3			Measles		
MATERNAL EDUCATION	**Male**	**Female**	***Significance (p-value)*** [a]	**Male**	**Female**	***Significance (p-value)*** [a]
None	19	13	*0.0506*	25	20	*0.4486*
Primary	37	34	*0.6466*	39	39	*0.8902*
Secondary	53	59	*0.0711*	48	76	*0.9682*
Significance (p-value) [b]	*0.0000*	*0.0000*		*0.0019*	*0.0000*	
Source: (*24*)						

Note: n/a indicates that the cell comprises fewer than 25 cases.
[a] The null hypothesis is that poor and non-poor are the same in each row.
[b] The null hypothesis is that within this stratifier (i.e., education, region or residence) and each column, all classes are the same.

Other indicators such as infant mortality were not included in the simultaneously-stratified tables, because the number of events (deaths) would be too few to construct robust rates. The full set of *p*-values for the simultaneously-stratified tables were beyond the scope of this work, although selective test statistics were run and were used to screen particular effects reported.

The simultaneously-stratified tables are analogous to the singly-stratified tables, with each class on the left broken further into subgroups based on the other stratifiers (see Table 7). Thus, in Table 7, the first two cells with values are the percentages of women with no education (poor, and not poor, respectively) who were attended by a skilled birth attendant (SBA). Table 7 is a small subset of a simultaneously stratified table. Note that even among women with no education, position relative to the poverty line makes a tremendous difference in access to SBA. In fact, for each level of education, each region and area of residence, further sub-stratification by ‘poverty line status’ results in wide disparities in access to an SBA. Significance tests would be needed for each pairing to determine whether these disparities are statistically significant (Table 8 in the next chapter shows that in fact all of the results from this simultaneous stratification are significant).

If we were to rely solely on the singly stratified analysis for policymaking, the Western region might be prioritized for improvements in access to a SBA because when comparing regions, this area had the lowest percent of women with access to a SBA (33 percent). However, in substratifying regions by poverty line, as in Table 7, it becomes clear that the poor in the Coast region have extremely low levels of SBA (14 percent) despite the region’s greater overall SBA rates.

Table 7. Skilled Birth Attendant in Kenya: Poverty Status Simultaneously Stratified

STRATIFIER 1		STRATIFIER 2		INDICATOR
Stratification Class	Stratification Level	Stratification Class	Stratification Level	Delivery Assisted by SBA
Education	None	Poverty Line	Poor	19
			Not Poor	40
	Primary	Poverty Line	Poor	24
			Not Poor	45
	Secondary or more	Poverty Line	Poor	43
			Not Poor	77
Region	Central	Poverty Line	Poor	n/a
			Not Poor	70
	Coast	Poverty Line	Poor	14
			Not Poor	49
	Eastern	Poverty Line	Poor	31
			Not Poor	56
	Nairobi	Poverty Line	Poor	n/a
			Not Poor	78
	Nyanza	Poverty Line	Poor	24
			Not Poor	52
	Rift Valley	Poverty Line	Poor	24
			Not Poor	50
	Western	Poverty Line	Poor	26
			Not Poor	39
Residence	Rural	Poverty Line	Poor	25
			Not Poor	49
	Urban	Poverty Line	Poor	(40)
			Not Poor	72

Note: n/a indicates that the cell comprises fewer than 25 cases. Parentheses ( ) indicate less than 50 cases.

Chapter 9: Reading and Visualizing Results from Simultaneous Stratification

As described above, 'simultaneous stratification' allows a more detailed analysis of inequalities in health. Such an analysis embraces the reality of multiple forms of social stratification that are at work in all communities in all countries. By examining gender and ethnicity *together*, for example, we can gain a better understanding of how two forms of social exclusion might interact. In some cases, being both female and a member of a disadvantaged ethnic group will confer greater health disadvantage than being a male in the disadvantaged ethnic group or being just female and of a more advantaged ethnic group. Interpreting the tables generated using simultaneous stratification is similar to interpreting the singly stratified tables. The following section has a few instructional examples. As with the singly-stratified tables, the tables lend themselves to answering fairly direct questions as well as more comprehensive analysis involving slightly more complex interpretation and visualization.

Simple Reading of Simultaneously-Stratified Tables

First, let's examine several direct questions that could be answered using the simultaneously stratified tables:

- ***Does educational level of the mother make a difference in whether there are gender differentials in vaccination of children?***

In Ethiopia, while a singly stratified relationship suggests a slight (though statistically insignificant) male advantage over female children for all vaccines, the more in-depth simultaneously stratified analysis indicates an interaction with education: sons of uneducated women have slightly better access to all immunizations measured but DPT1, which is given at rates that are indistinguishable. Sons and daughters of primary-educated mothers have more or less equitable access to immunization while daughters of mothers with secondary or more schooling have significantly higher measles vaccinations than do sons (see Table 7 in the previous chapter). Unlike the situation in many other countries, it appears that basic immunization is very inequitably distributed, suggesting that there are significant challenges in the current implementation of even vertical programs.

- ***How is access to a Skilled Birth attendant (SBA) affected by poverty status and educational level in Kenya?***

As Table 8 indicates, access to an SBA is stratified by level of education amongst the 'not poor.' Note that all of the differences highlighted here are highly statistically significant as evidenced by the *p*-values presented in italics. Non-poor women with no education use SBAs 40 percent of the time while women with primary education have slightly higher access at 45 percent and women with the highest educational levels access a SBA for 77 percent of deliveries. For the poor, on the other hand, only 19 percent women with no education access a SBA while women with primary and secondary or more education access SBA at rates of 24 percent and 43 percent, respectively. Importantly, the most highly educated poor women have access to SBA at a rate only equivalent to the rates attained by the non-poor with no education. Thus, in access to SBA, education is helpful, but still dramatically constrained by poverty status.

Table 8. Skilled Birth Attendant in Kenya: Simultaneous Stratification with *p*-values

		POVERTY LINE		
Stratifier Class	**Level**	Not Poor	Poor	*Significance (p-value)*[a]
Education	None	40	19	*0.0000*
	Primary	45	24	*0.0000*
	Secondary or more	77	43	*0.0000*
	Significance (p-value)[b]	*0.0000*	*0.0000*	
Region	Central	70	n/a	*n/a*
	Coast	49	14	*0.0000*
	Eastern	56	31	*0.0000*
	Nairobi	78	n/a	*n/a*
	Nyanza	52	24	*0.0000*
	Rift Valley	50	24	*0.0000*
	Western	39	26	*0.0003*
	Significance (p-value)[b]	*0.0000*	*0.0003*	
Residence	Rural	49	25	*0.0000*
	Urban	72	40	*0.0054*
	Significance (p-value)[b]	*0.0000*	*0.0057*	

Note: n/a indicates that the cell comprises fewer than 25 cases.
[a] The null hypothesis is that poor and non-poor are the same in each row.
[b] The null hypothesis is that within this stratifier (i.e., education, region or residence) and each column, all classes are the same.

More Complex Visualization of Simultaneously-Stratified Tables

Rather than asking one narrow question, we may wish to take a particular indicator, like ‘underweight’ for children and do an analysis across several different stratifiers to determine the varying impact of social stratifiers on outcomes. For example, in Ghana, the percentage of underweight children varies by ethnic group, educational level of the mother, region, and residence.

A brief overview of all the results for ‘underweight’ reveals some striking differentials between urban and rural dwellers (see Table 9). We look first at stratification within educational level of the mother—those in rural areas with no education have underweight of 32 percent, for primary education it is 30 percent and secondary education it’s 24 percent. By contrast, in urban areas, these rates are 19, 17 and 14 percent respectively. Amongst rural dwellers as a whole, the educational differences are highly significant stratifiers of underweight (*p*-value of 0.0001), while amongst urban dwellers the educational differences are not significant (*p*-value of 0.3471). Looking at the test statistics in the other direction, each row is significant at the 95 percent confidence level meaning that for each level of education, the difference between urban and rural is significant.

We turn next to ethnic group—and note that the differences between ethnic groups, when stratified by urban and rural residence, are dwarfed by the obvious differences between living in an urban area (14-19 percent underweight) and a rural area (27-31 percent underweight). Note also that the differences between educational levels are not significant for either the rural group or the urban group. The test statistics the other way reveal that the urban/rural differences within each educational group are highly significant.

Finally, we might look at the intersection between region and residence to determine whether a single region is 'responsible' for these differences or whether the effect persists in many regions. We can see in Table 9 that there are four regions which have enough of an urban population to generate a statistically significant figure and that there are nine regions which have a sizeable rural population. All of the urban groups have underweight below 20 percent whereas all the rural groups, across nine different regions have underweight percentages above 20—with the Upper East and Northern regions at 35 and 41 percent, respectively. Note that the regional differences amongst those who are urban are not statistically significant (*p*-value is 0.3191) but that the regional differences amongst those who are rural are highly significant (0.0012). And looking at the test statistics in the other direction, it's clear that the urban/rural differences are significant only in the following four regions: Ashanti, Eastern, Northern, Western.

In order to link this analysis to policy, one might wish to go further, comparing other child health outcomes to these. For example, undernutrition is, of course, a health issue but may be remediated through different policies and pathways than would differences in immunization. In addition, because a large proportion of U5MR is linked to malnutrition, it would be interesting to examine whether the inequities here are similar to those found for 'undernutrition.' This can be done using the simultaneously stratified tables (though in the case of Ghana, we do not have IMR, NNMR and U5MR). In addition, we may want to compare another child health outcome like DPT3 to the results for underweight. For example, underweight in urban areas is 16 percent for girls for male and female, but for DPT3, females have higher coverage than male (87 versus 81 percent). In rural areas, though males have higher rates of underweight, females are less likely to have DPT3 (68 for females, 70 percent for males).

Table 9. Underweight in Ghana Stratified by Region, Residence, Ethnicity Grouping and Education

STRATIFIER			UNDERWEIGHT (%)	
Class	**Level**	**Significance** *(p-value)*	**Rural**	**Urban**
Education	None	*0.0030*	32	19
	Primary	*0.0076*	30	17
	Secondary or more	*0.0028*	22	14
Significance *(p-value)*			*0.0001*	*0.3471*
Ethnicity Groups	Dominant-primary	*0.0026*	31	14
	Dominant-secondary	*0.0001*	27	14
	Not dominant	*0.0016*	27	18
Significance *(p-value)*			*0.4355*	*0.4448*
Region	Ashanti	*0.0206*	28	15
	Brong Ahafo	*0.8329*	24	(23)
	Central Region	*0.1645*	29	19
	Eastern Region	*0.0186*	25	13
	Greater Accra	*0.6761*	(10)	13
	Northern Region	*0.0381*	41	(24)
	Upper East Region	*0.1491*	35	n/a
	Upper West Region	*0.2168*	27	n/a
	Volta Region	*0.9221*	25	n/a
	Western Region	*0.0198*	28	(12)
Significance			*0.0012*	*0.3191*

Note: n/a indicates that the cell comprises fewer than 25 cases. Parentheses () indicate that the cell comprises less than 50 cases.

[a] The null hypothesis is that poor and non-poor are the same in each row.

[b] The null hypothesis is that within this stratifier (i.e., education, region or residence) and each column, all classes are the same.

Figure 14 is an elaboration of Figure 12 in Chapter 7. This Figure shows how simultaneous stratification can be portrayed on a spider graph. In Figure 14, each ‘spoke’ of the graph represents two forms of stratification—sex and ethnicity (i.e. simultaneous stratification). Importantly, the simultaneous stratification by sex and ethnicity together resulted in more pronounced differences between males and females. In this spider graph we see that the rate of ‘moderate underweight’ amongst children is highest for the simultaneously stratified subgroup ‘male’/ ‘not dominant ethnic group’. Males fare worse than females in the not-dominant group, while males and females to have more or less equal rates of underweight in the dominant ethnic group.

Figure 14. Spider Graph of Moderate Underweight with Simultaneous Stratification, Cambodia

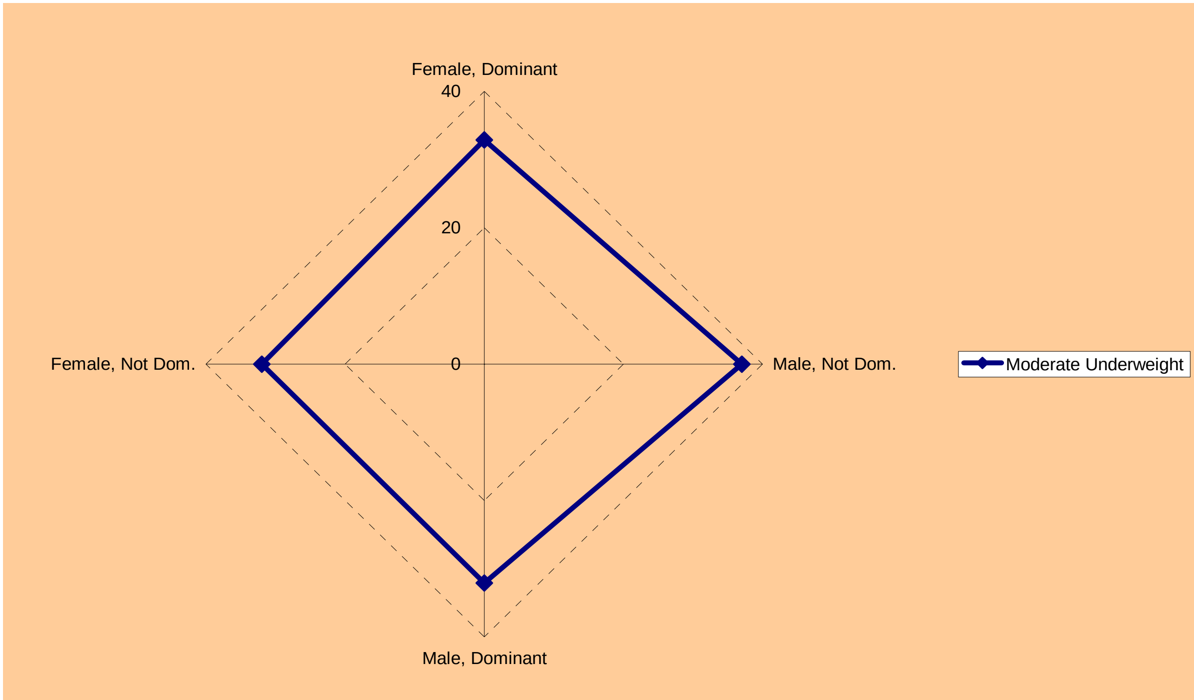
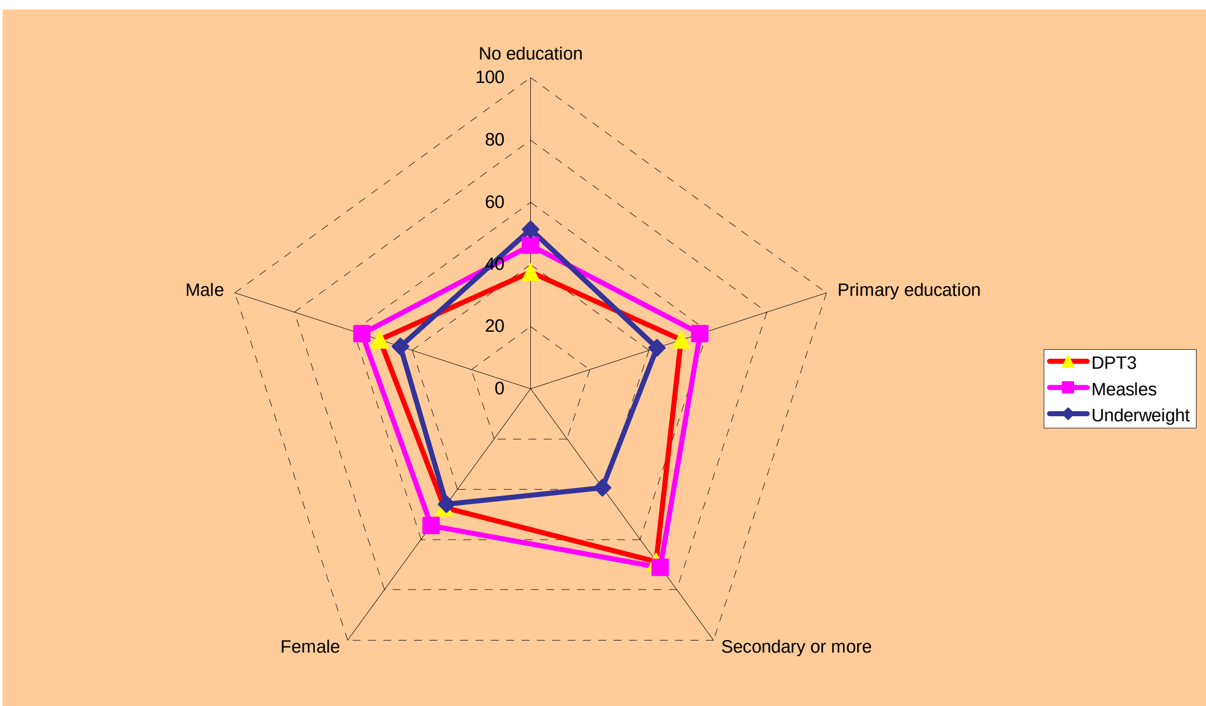
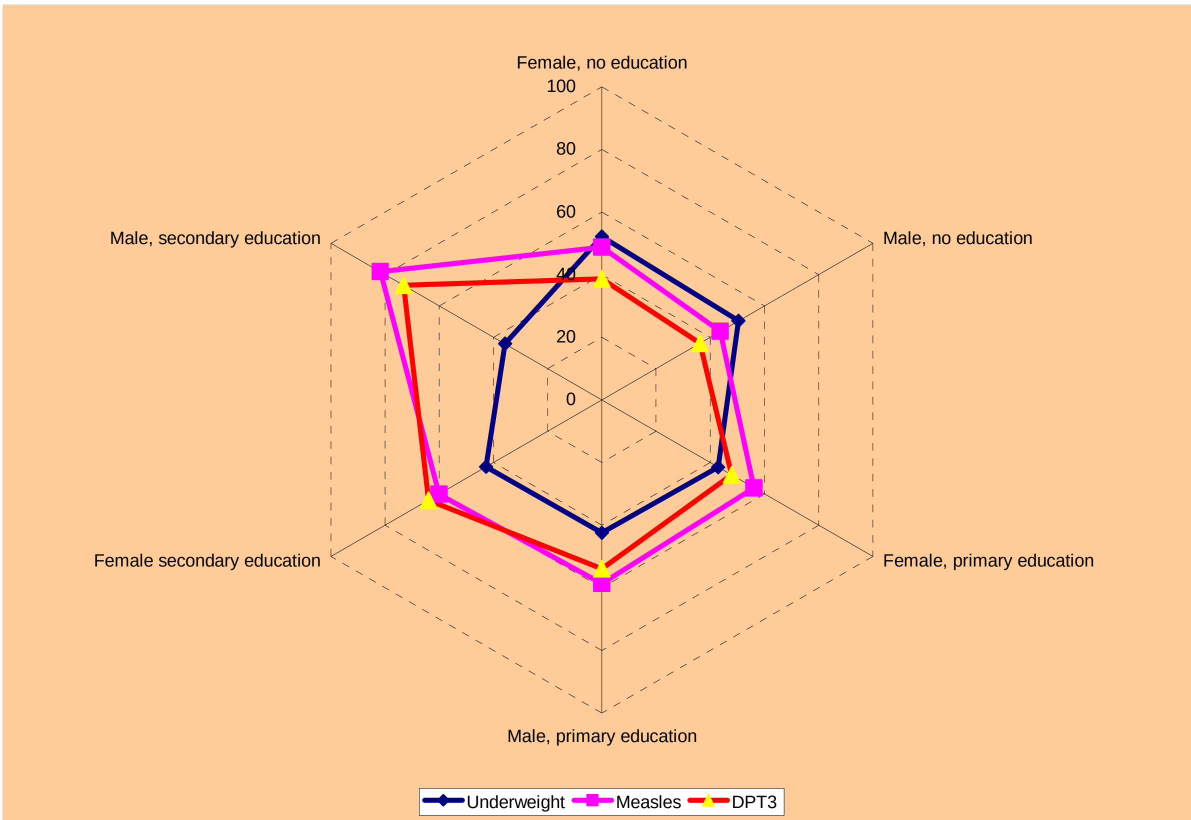


Figure 15. Spider Graph of Child Health Indicators Stratified by Sex and Maternal Education, Cambodia



Figures 15 and 16 are presented as a comparison. Figure 15 is based upon singly-stratified data while Figure 16 takes advantage of simultaneously-stratified data. Figure 16 is a spider graph used to display three types of child health indicators (the strands of the web) across two types of stratifiers (sex and maternal education). While this figure makes the advantage of secondary education clear, it is not possible to see the combined effect of sex and maternal education. By contrast, the spokes of the web in Figure 16 represent two stratifiers each, depicting the combined effect of sex and maternal education. In this graph, male children with mothers who have a secondary education are clearly advantaged in terms of higher immunization and lower rates of underweight as compared to all other groups depicted. In fact, females with mothers who have secondary education appear to have rates of underweight very close to the female/primary education and male/primary education groups. A similar pattern exists for measles immunization.

Figure 16. Spider Graph of Child Health Indicators Using Simultaneous Stratification, Cambodia



Chapter 10
Interpreting Tabular Results with Maps

Spatial data can be a helpful tool in visualizing inequity, especially when regional differences are large. In addition, particularly in areas with representative governments, maps can be a useful political tool, highlighting regions or subregions that are falling behind. Maps can be used to show singly-stratified relationships (for example, percentage vaccinated against measles by region), which are simple and easy to understand. Geospatial representation can also be used to show simultaneous stratification by using multiple maps. For example, one could make an urban map of measles vaccinations by region and a rural map of measles vaccinations by region.

One can use large regions or smaller subdivisions, but it is important to note the advantages and shortcomings of each method. Large regions tend to mask heterogeneity within a region by aggregating data from subsections and providing only an average. While coarse geographic categories have the problem of aggregation, the catch with fine divisions is that sample sizes can become too small to provide reliable data. The sample of people with secondary education in certain regions may be too small to determine if there are differences by sex or gender. Furthermore, other things equal the 'smaller the unit of grouping, the larger the observed inequality' (Anand et al 2001).

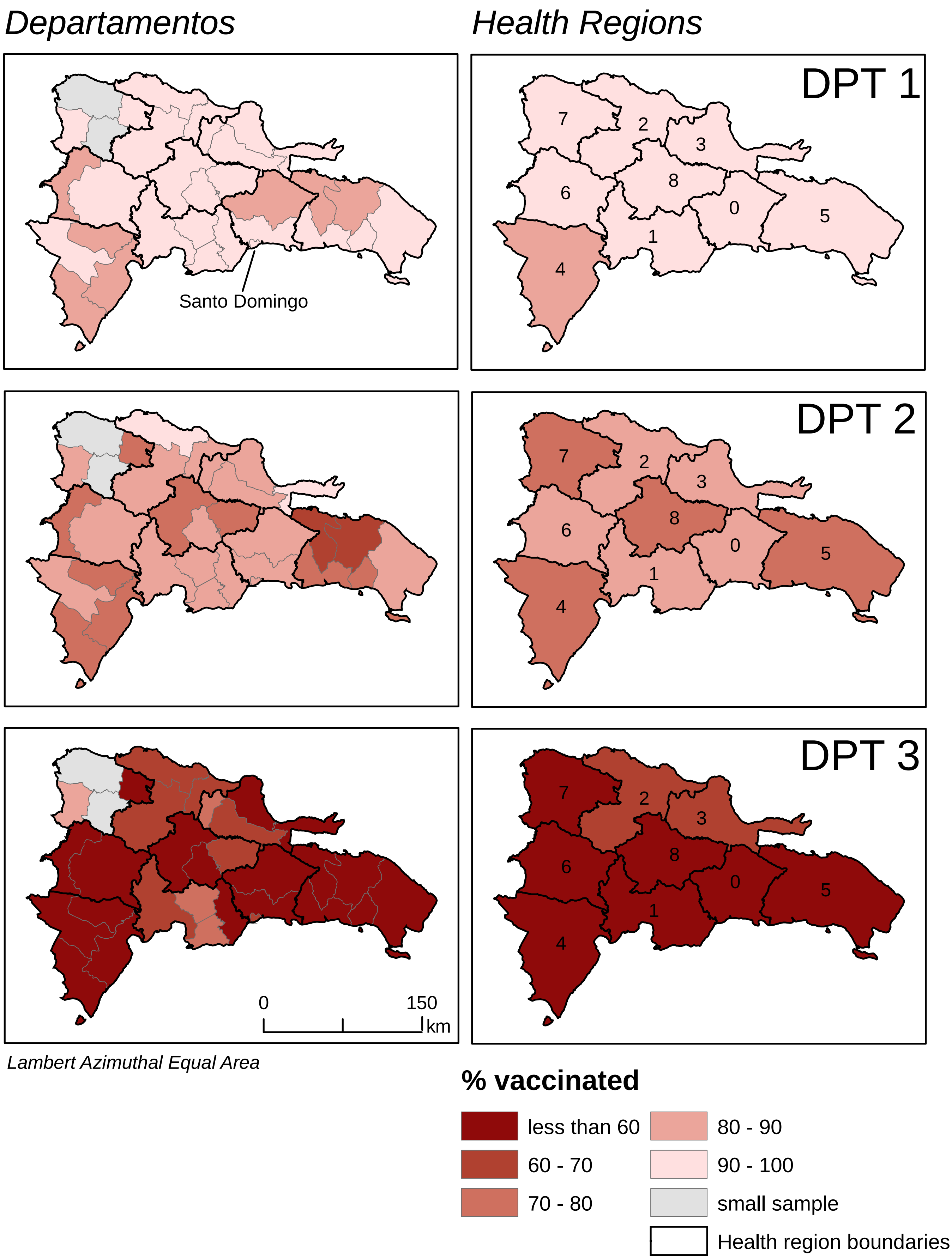
In translating tabular data into maps, there are a few important issues to note. Spatial data has to match, including names and spellings of regions (for example, San Jose and San José do not match and numerical data for that region will not appear on the map). Thus, the data must be checked carefully before being exported to a mapping program. In this chapter, we look at two examples of maps used to highlight equity data.

DPT coverage in Dominican Republic

The Dominican Republic provides an excellent example of the need for disaggregated data (Figure 17). There are 32 provinces, grouped into eight 'health administrative regions.' Within each of the administrative boundaries, one can see the amount of heterogeneity that would be masked by using only the eight larger divisions. The Dominican Republic also has a fairly large DHS survey, which means that most sample sizes are large enough to be reliable, even with 32 regions.

While a single indicator may give insight into the overall health of a population, it does not always capture the full health situation. Vaccination campaigns can immunize thousands of children against measles, but some argue that this is not a sustainable method of delivering *all* health care services. Examining the DPT sequence shows who is receiving consistent care and which parts of the population 'drop off' and never receive boosters or subsequent doses. In the example from the Dominican Republic, one can see that in the majority of the regions, over 90 percent of the population of children under-one year received the first DPT shot. None of the regions have rates that are below 50 percent for DPT1. For the second DPT shot, every region except Puerto Plata and Samana have vaccination rates between 50 and 90 percent. By the time children need the third DPT shot, not a single region has over 90 percent vaccinated and five regions emerge where the vaccination rate is below 50 percent. The examination of DPT vaccines illustrates how geospatial data can be used to determine health care access. In this case, analysis of a single health indicator may be used as a proxy for who has access to well-functioning or sustainable health systems.

Figure 17 (following page). Map of DPT vaccination rates, Dominican Republic



Departamentos
Health Regions
DPT 1
DPT 2
DPT 3
Santo Domingo
7
2
3
6
8
0
5
1
4
0
150
km
Lambert Azimuthal Equal Area
% vaccinated
less than 60
60 - 70
70 - 80
80 - 90
90 - 100
small sample
Health region boundaries

AIDS Knowledge in Cambodia

The distribution of AIDS knowledge (specifically, the percentage of women who know that 'using a condom can prevent AIDS') can be measured in Cambodia using spatial data for 17 regions. The maps in Figure 18 stratify the AIDS knowledge indicator by both region and by level of education. The first map in the upper left is a singly-stratified map simply showing the results for all women but across the 17 regions. This map makes clear the heterogeneity in the distribution of AIDS knowledge across the regions in Cambodia.

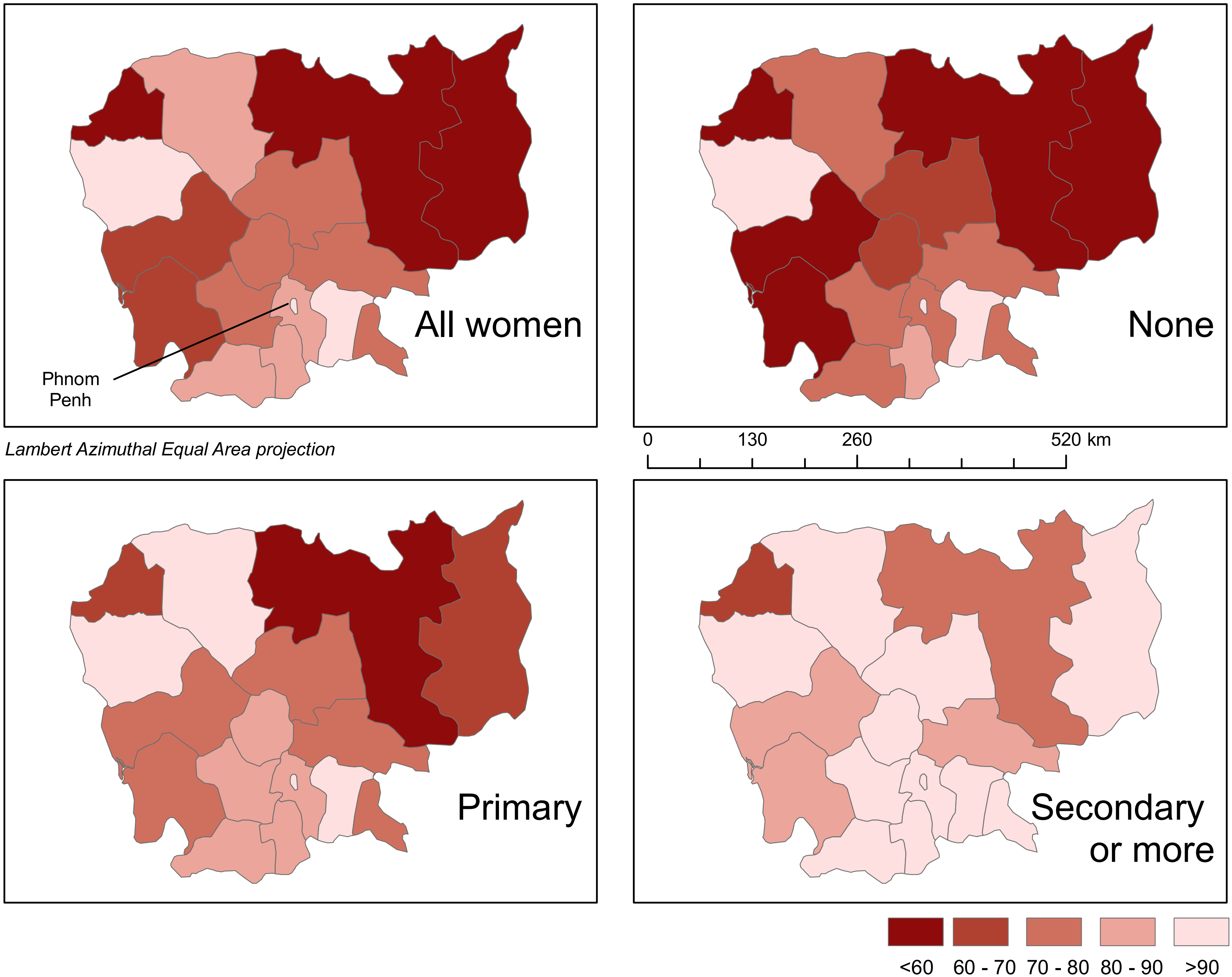
The other three maps each represent a level of education thereby simultaneously stratifying the data by region and by level of education. What is immediately apparent in the maps is that two large regions in the northeast and one in the northwest consistently lag behind the rest of the country with rates of AIDS knowledge below 70 percent for this particular indicator for women with no education and women with primary education. By contrast, three regions (including one that borders the aforementioned northeastern region) have AIDS knowledge above 90 percent even among women with primary education or less.

Women with secondary education are clearly at a marked advantage in terms of AIDS knowledge, though in the one northwestern region that is still dark in color, education still does not appear to guarantee greater knowledge in this area.

The maps are useful for displaying heterogeneity in a powerful visual format. In addition, they allow for analysis of whether particular geographic regions or corridors of the country may be isolated or falling behind. Of course, further verification of the statistical significance of these differences needs to be undertaken to verify these potential differences.

The maps of Cambodia demonstrate considerable heterogeneity across regions. In some areas of the country, diffusion of knowledge is such that level of education does not appear to make a difference in levels of AIDS knowledge. In other areas, we can see that with each level of education the percentage of women with AIDS knowledge about condoms increases. Although the first map of all women might lead us to believe that geography plays a major role in stratifying outcomes, the final map complicates conclusions. Clearly, region appears to be a stratifier, but the fourth map displays a far more equitable picture than the others, implying that level of education is a powerful equalizer of AIDS knowledge. For women with a secondary level of education, region no longer matters as much for their level of AIDS knowledge. Further research would be necessary to determine the reasons for these differences as well as the policy choices available for improving access to AIDS knowledge amongst those with lower levels of education.

Figure 18 (following page). Map of AIDS Knowledge (percentage of women who know that using a condom can prevent AIDS), by level of education, Cambodia



All women
Phnom
Penh
None
Lambert Azimuthal Equal Area projection
0
130
260
520 km
Primary
Secondary
or more
<60
60 - 70
70 - 80
80 - 90
>90

PART III
Chapter 11
Monitoring Trends over Time

Once an equity baseline is set, the important task of monitoring trends over time begins. Though a considerable literature on identifying health gaps has been developed, the experience in monitoring health gaps is less well documented.

This section provides a brief overview of some technical and conceptual considerations for describing changes in inequality. The section begins with a discussion of how to determine whether disparity in a particular indicator increases or decreases. This section is followed by a discussion of how to analyze how an indicator's average compares to its distribution over time.

Statistical Analysis of changes in Disparity

In a trend analysis, the crucial question is whether a summary indicator of disparity changes through time. The analysis needed to answer this question involves two steps.

First, from the single and simultaneous stratification described above we have learned how to determine which groups show different and statistically significant outcomes from one another in a given year. The ratio of these outcomes, which can be called the 'rate ratio', will generally be different for two points in time.

Table 10. Statistical Analysis of Changes in the Rate Ratio

	YEAR 1	YEAR 2
Group A	GA_1	GA_2
Group B	GB_1	GB_2
t-test	GA_1 and GB_1 are significantly different	GA_2 and GB_2 are significantly different
Ratio (relative gap)	GA_1/GB_1	GA_2/GB_2

Second, we must test *the statistical significance of observed differences of rate ratios* between two time periods is statistically significant. Put in terms of the shorthand used in Table 10, we must determine whether GA_1/GB_1 and GA_2/GB_2 are statistically different from one another.

This is process is best described through an example. We will use one example based upon the top and bottom wealth quintiles for U5MR in two rounds of a survey and a second looking at maternal education and rural/urban dwelling for measles vaccination. In the first survey round, we find the following two results:

a) The U5MR for the top quintile is 30 and the U5MR for the bottom quintile is 90, meaning that the rate ratio is 3

b) Mothers who are urban and have higher education have children with measles coverage of 80 percent whereas those who are rural and have no schooling have children with measles coverage of 60 percent, yielding a rate ratio of 1.3

Notice that these ratios imply that the poorest children are three times more likely to die than richest children, and that urban, highly educated mothers are 30 percent more likely to immunize their children against measles. In one case (measles) the better-off group is placed in the numerator of the ratio and in the other case (U5MR) it is the worse off who are placed on the numerator. Although this is not the only way to express the ratio, it is the most intuitively appealing and therefore a convention of convenience. For comparison, it is of utmost importance to maintain whatever convention is used in a consistent fashion across time.

For the second round, we observe the following:

a) The U5MR for the top quintile is now 25, the U5MR for the bottom quintile is 80, resulting in a rate ratio of 3.2

b) Mothers who are urban and have attained higher education have children with measles coverage at 84 percent whereas those who are rural and have no schooling have children with measles coverage at 70 percent, making the rate ratio 1.2

We then ask, 'Has the rate ratio in U5MR increased (wealth disparity increased)? Has it declined for measles coverage (in terms of residence and education)?' The rate ratios for U5MR are 3 and 3.2 respectively in the two rounds of the survey are 3 and 3.2 for the U5MR example and 1.3 and 1.2 for the measles example. Are these rate ratios statistically different or do they just represent insignificant minor variations?

A simple approach is to construct a test of differences for the components of the ratios. In other words, to find out if GA_1 is statistically different from GA_2 and GB_1 is statistically different from GB_2. In terms of the examples above, is an estimated U5MR of 30 different from one of 25, given the margins of error in these estimates? And are 80 and 90 different? Similarly, is a measles immunization rate of 80 statistically different from one of 84? And are the estimated immunization rates of 60 and 70 different?

When using this test of differences, it is helpful to first classify cases according to the pattern that leads to the increase, decrease or lack of change in disparity[20]. Tables 11a and 11b depict different scenarios based upon what might occur when the indicator for group A is better than for group B in the first round. Increases in disparity between survey rounds are possible in each of the cases displayed in Table 11a. For example, disparity might increase if both Group A and B improve, but B improves less than A, or, disparity might increase if Group A improves but Group B remains constant or declines. Likewise, several scenarios for a decrease in disparity are displayed in Table 11b.

[20] In this case we are constructing a new variable (the difference of the U5MRs or measles immunization) and estimating a new variance which results from the addition of the two estimated variances.

Table 11a. Cases of Increases in Disparity					
	CASE 1	CASE 2	CASE 3	CASE 4	CASE 5
Group A	improves	improves	constant	improves	declines
Group B	constant	declines	declines	improves (less than A)	declines (more than A)

Table 11b. Cases of Decreases in Disparity					
	CASE 1	CASE 2	CASE 3	CASE 4	CASE 5
Group A	constant	declines	declines	declines	improves
Group B	improves	constant	improves	declines (less than A)	improves (more than A)

In this case, 'constant' means that the null hypothesis Group A indicator (first round) = Group A (second round) was not rejected.[21]

Table 12 shows how the t-test could be applied to cases where disparity seems to have increased. It is important to stress 'seems', as this is the hypothesis that needs to be tested. The table deals with an increase in wealth disparities in U5MR (following the example above in Table 11a). Thus, when the indicator for the worse-off group (in this case the bottom quintile, Q1) declines, it is an improvement. If the example referred to knowledge about HIV/AIDS, an increase would be an improvement.

If both groups show constant results, obviously there was no change in the ratio (see Case 1). If the situation of the better off group improved while that of the worse off deteriorated, obviously disparity increased. If the indicator for the worse-off group was constant (or declined) and it improved (or was constant) for the other one, then disparity increased (cases 2 and 3 of Table 12). If the indicator for both groups moved in the same direction, whether disparity increased or decreased depends on the relative rate of change in each group. The U5MR example falls into the category of Case 5—the U5MR decreased for both the top and bottom quintile.

[21] If it is assumed the variances are different in each sample, the joint variance can be calculated by adding the individual ones. For this and all other tests described below, the significance level is traditionally set at 5 percent. Because the DHS and MICS, like many other surveys are cluster surveys, a specific formula is required to estimate the variances.

Table 12. Possible t-test Results by Groups When Testing for Increase in Wealth Disparity, U5MR

CASE 1	CASE 2	CASE 3	CASE 4	CASE 5
Q1 constant Q5 constant	Q1 increased Q5 decreased	Q1 constant (or increased) Q5 decreased (or constant)	Q1 increased Q5 increased	Q1 decreased Q5 decreased
Constant disparity	Increased disparity	Increased disparity	Unclear trend in disparity	Unclear trend in disparity

Even when differences are significant for both quintiles and even if the percentage reductions are almost twice as large for one group as compared to another, a further test of significance is needed to determine whether the changes are of the same magnitude.[22] This can be done by using a one-tailed hypothesis test that the indicator was reduced by a certain percentage, for example 50 percent (i.e. in the first period it was double the level in the second period) for each group. The test can be performed for 25 percent or other values (depending on the magnitude of the observed changes for the different groups). For instance, in the U5MR case, the top quintile was reduced by 17 percent (from 30 to 25) and for the bottom quintile it was reduced by 11 percent (from 90 to 80). Then, tests can be performed to see if the changes for each group exceed 10, 15, and 20 percent.

If the hypothesis of no change is rejected for one group but it is not rejected for the other group there is clear indication of an increase in disparity. For instance, if the hypotheses of 20 percent and 15 percent reduction are not rejected for the wealthiest group but it is rejected for the poorest group, we have reason to believe that disparities have increased.

Table 13 shows a similar analysis for the case when disparities seem to have declined. The example of simultaneous stratification for measles immunization presented above is used. If both groups -urban and educated (UE) and rural with no schooling (RNS) for short- maintained the same level of immunization coverage (i.e. neither $UE_1 – UE_2$, nor $RNS_1 = RNS_2$ are rejected), then disparity has remained constant. If according to the t-tests the indicator for the worse-off group was constant (or declined) and it improved (or was constant) for the other one, then disparity increased. If the indicator for both groups moved in the same direction (i.e. tests show that both EU and RNS are different for rounds one and two), whether disparity increased or decreased depends on the relative rate of change in each quintile and the one-tailed tests as described above would be needed.

[22] For a poignant discussion of the misguided reliance on strict application of t-tests in economics, without taking into account the 'broad' picture the data present, McCloskey and Ziliak (1996) can be consulted.

Table 13. Possible t-test Results by Groups When Testing for Reduction Educational Disparities				
CASE 1	**CASE 2**	**CASE 3**	**CASE 4**	**CASE 5**
UE constant RNS constant	UE increased RNS decreased	UE constant (or increased) RNS decreased (or constant)	UE increased RNS increased	UE decreased RNS decreased
Constant disparity	Increased disparity	Increased disparity	Unclear trend in disparity	Unclear trend in disparity

In the measles example with which we started, among urban, highly education mothers children's vaccination went from 80 to 84 percent (an increase of 5 percent) over the two survey rounds while the rural uneducated mothers' coverage went from 60 to 70 percent (an increase of 17 percent). If both of these changes are different from 0, this places this example in the 'case 4' column. Then we need to test if they are also different from 10, 15 and 20 percent. If both changes are statistically different from a 20 percent change but not from a 15 percent change, than we should conclude that disparities have not changed.

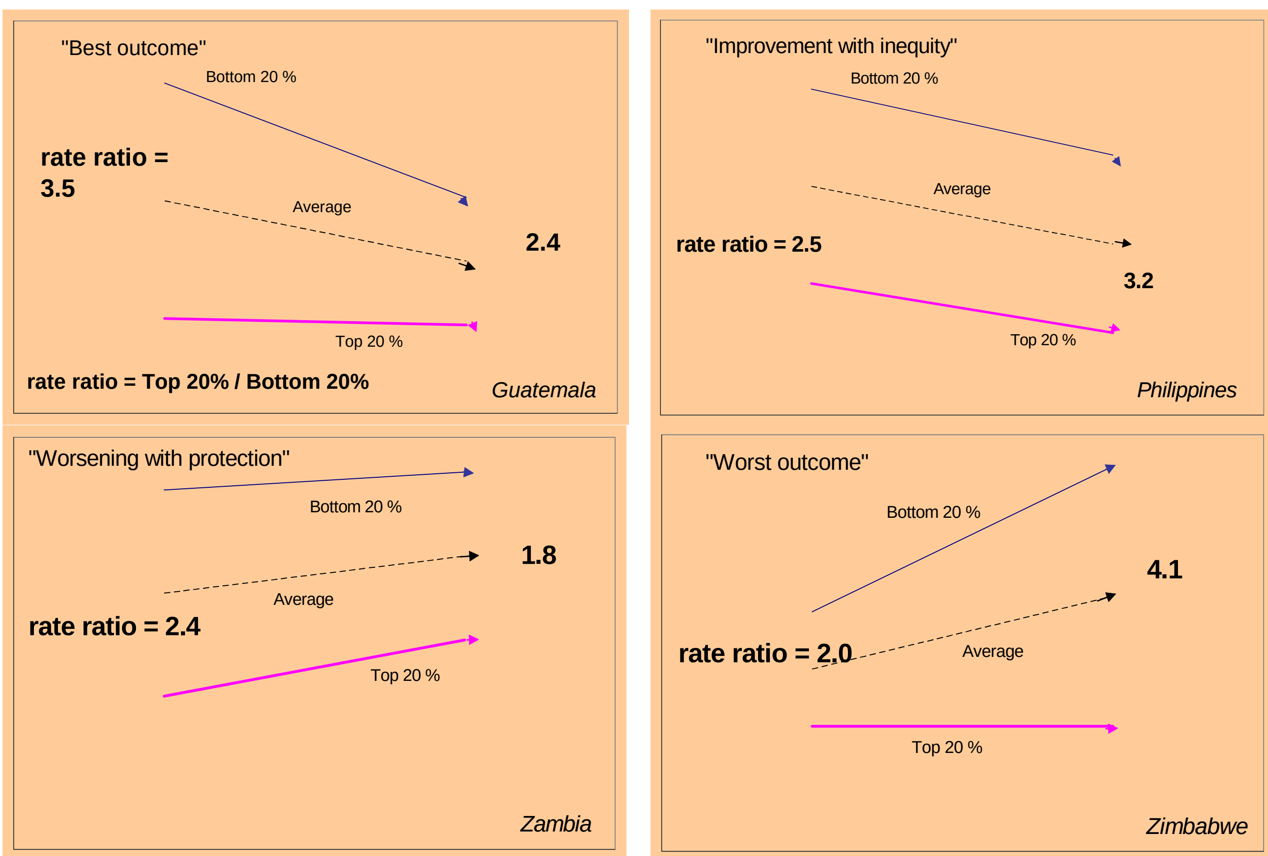
Relationship between average levels and disparity

In this section we look at an indicator's 'average,' for example, the country average, and compare its movement over time to the changes in 'disparity' (also called the 'distribution'), the difference between population groups of choice. The relationship between changes in an indicator's average and in the ratio between population groups is not a simple one. A change in the average could result from any number of changes in its distribution (and consequently in disparity). Thus, changes in the national average do not imply *a priori* any particular modification in equity. For example, an improvement in U5MR could go together with increases in the ratio which might be a result of a worsening situation among disadvantaged and vulnerable children. Combining the various possibilities between improvements and deterioration in the average and in its distribution, we obtain the four alternative scenarios described in Table 14. Although not all of them are equally likely to happen, Figure 19 shows an example of each for U5MR trends stratified by wealth quintiles.

Table 14. Possible Relationships Between Changes in Average and Disparity

TRENDS		DISPARITY (rate ratio)	
		Narrowing	Widening
AVERAGE	Improving	*Best outcome*	Improvement for better-off children, but not for the disadvantaged
	Worsening	Worsening with an element of protection of the disadvantaged	*Worst outcome*

Figure 19. Scenarios for Changes in U5MR Over Time: Average Level and Disparity Between Top and Bottom Wealth Quintiles



Note: These figures present some country examples from surveys in the mid/late 1980s and mid 1990s and are based on a dataset not used in the rest of the *Guide*. (Minujin and Delamonica (2003))

Table 15. Scenarios for Changes in Average U5MR and Rate Ratio

TRENDS		DISPARITY (rate ratio)		
		Narrowing	Constant	Widening
Average	Improving	*Best outcome*		
	Constant			
	Worsening			*Worst outcome*

Given that the analysis of the statistical significance of the differences should also be carried out for the average level, Table 13 can be modified to look like Table 14 as there are situations in which no changes are observed (either in ratios, average levels, or both). The table can be filled with the results of the trend analyses for different indicators and stratifiers.

Time interval between surveys

An issue that is likely to arise in many countries, especially in the future as more surveys are undertaken, is that testing differences may depend on the intervals between the surveys. For instance, if six years elapse from one round to a second round but only three years separate the second round from the third one), the conclusions could be affected. Changes may only be significant after long periods of time, but not when surveys are separated by a few years, it is worthwhile to explore the connection between the intervening years and the statistical significance of the observed changes.

An analysis by Minujin and Delamonica (2003) revealed that the length of the time period between the surveys usually does not undermine the conclusions regarding trends in inequality. However, it does seem to be the case that at least four years are needed in between rounds to obtain significant changes. Although this might vary from country to country and indicator to indicator, it is worth remembering when establishing a monitoring plan (especially given the cost of surveys).

Changes in ranking between surveys

The inter-temporal comparisons of disparities for the province/state (geographic) or other stratifiers should be carried out very carefully, as in some cases the worst or the best province is not the same for different rounds, and thus the ratio is not strictly comparable. This issue is similar to the one mentioned above in terms of the convention of which value to place in the numerator. This problem does not emerge when analyzing quintiles, as the bottom one is always the same quintile. Also, quintiles always represent 20 percent of the population, while other stratifiers may result in groups of heterogeneous sizes in different survey rounds.

In contrast, the elements that make up the wealth index stratifier often vary between survey rounds. Box 2 contains an explanation of approaches that may be taken to deal with this issue.

BOX 2. Challenges of Using the Wealth Index to Monitor Trends

The wealth index is based upon questions in the survey that address the assets held by a household. Because the questionnaire evolves over time, the set of asset indicators assessed may change between survey years, presenting a challenge for analysis of trends over time. There are two approaches one could use to address this.

First, for each of the available years, one could use a different set of asset indicators. Although there might be differences in the components of the wealth index, in each year all the information on assets available in the survey would be used. The drawback is, of course, that the basis of the classification of households is different, amplifying and confusing the source of the observed variations.

Alternatively, one could attempt to use a consistent subset of indicators to compare data in two surveys. The 'consistent at the end' approach would start with the set of asset indicators used in the latest survey round and when analyzing the earlier rounds include only these asset indicators which occur in the last round. Similarly, the 'consistent at the beginning' approach would use the set of first round asset indicators as the standard for the wealth index, such that in later rounds of the survey only those assets which also appear in the first round are used.

Finally, in order to maximize comparability through time only the assets which are common across surveys can be used. This requires that the common assets be the ones found in all rounds. This, of course, implies some loss of information (efficiency) in the estimates. It is important to measure this loss empirically, but it seems that the benefits in terms of comparability and length of period under analysis compensate this loss. Minujin and Delamonica (2003) tested this by looking at the effect of deleting a few asset indicators from the construction of the wealth index in over 20 countries. The rank correlation of households using all or some sub-sets of the available assets averaged above 0.9. In other words, households were ranked using the wealth index calculated with all available information; then households were ranked again but using the wealth index with the smaller set of indicators (i.e. the ones common to both periods). There was very little difference in the rankings (almost 90 percent of the households were in the same position/rank).

Analyzing top versus bottom or across the whole distribution

Finally, a few words on whether only some comparisons should be carried out (i.e. the extremes of the distribution: top versus bottom quintile, or ethnic group with best results versus the one with the worst outcomes; called 'simple range measures') or the whole distribution should be included (Anand et al. 2001). Although theoretically it seems the latter is approach seems preferable, as it uses and conveys more information, there are a couple of practical problems. First, the tests are more elaborate and often result in inconclusive results, even if palpable disparities do change. Second, it is considerably easier to express the results in terms of simple ratios between two groups, especially to policymakers and other non-technical stakeholders.

Chapter 12
What have we learned? *Procedural Findings and Policy Implications*

This *Guide* began with the simple assertion that disparities in health exist even in poor countries; and furthermore, that the data currently exists for simple equity monitoring of the MDGs in these countries. The bulk of the *Guide* is devoted to technical instruction on how to operationalize equity-sensitive monitoring of the MDGs by disaggregating DHS and MICS data into population groupings and testing for statistical significance. Overviews of visualization techniques such as the use of spider graphs, whisker plots and maps are also included. However, monitoring is only useful if it results in concrete action to remediate disparities. A discussion of the full menu of policy choices that might evolve from such an analysis is outside the realm of this document, but some initial hypotheses about the links between the type of findings presented here and the next policy-relevant steps a country might take is warranted.

National averages are limited and mask a complex reality. In some places average health statistics are improving, but the gap between the most and least advantaged groups in society are quite large and increasing. Importantly, the majority of the social stratifiers tested in the examples used here yielded statistically significant differences across a wide range of indicators in six different countries. Thus, the *Guide* validates the assertion that stratification of a population into many groups can be fruitful in teasing out variation between groups that may not be obvious from the outset.

As expected, there were large differentials between geographic regions, and between urban and rural populations. Wealth quintiles (and status above or below the poverty line) were a strong stratifier of most health indicators. Differentials between male and female children were less marked in the single stratification but were more pronounced in some of the simultaneous stratification. In addition, educational attainment of mothers is a critical social determinant of most health indicators. Investments in education must be seen as having a dual positive effect in both the education and health sectors. Simultaneously, health messages and programs should be designed to reach less educated mothers and their children. And ethnicity, a core form of marginalization, remains understudied in the health and development literature and yet the examples presented here indicate often extreme ethnic differentials in key health indicators.

Equity analyses reveal that marginalization in health is multi-dimensional. Those who suffer from income poverty are also likely to be disadvantaged in levels of education, access to services and may also face ethnic discrimination. These inequities in health are interactive. Being disadvantaged in one area may lead to disadvantage in another area and causation may be blurred by feedback cycles. Simultaneous stratification is used in this *Guide* to assess the combined effect of two stratifiers. In multiple cases, disparities are small or nonexistent in single stratification—but more dramatic differences are uncovered using simultaneous stratification. In other cases, disparities across one stratifier, such as maternal education, are wide and yet sub-stratification by, for example, poverty status also yields significant gaps showing that poor women in all educational groups are far worse off than non-poor women. These results suggest that indeed the picture of health inequity is likely to be very complex in poor countries, and that policy measures need to be prepared to think strategically across the many dimension of poverty. In-depth analysis is warranted to tease out the underlying determinants of health disparities.

With a long list of indicators and multiple stratifiers, the simultaneously stratified tables can quickly result in a large amount of data that is a bit difficult to analyze. The use of pivot tables in Excel assists in making the data less unwieldy. Based upon pivot tables, simpler, more legible presentations of the data in the form of whisker plots, spider graphs and maps will help drive home the most salient inequities to key stakeholders.

The results presented in this *Guide* confirm the hypothesis that a sole focus on wealth quintiles will obscure other policy-relevant inequities. Regions are often coterminous with ethnic divisions or poverty profiles, although this codetermination is only revealed by simultaneous stratification. For example, measles vaccination rates seem to vary considerably by wealth, but when regions are added as substrata it becomes clear that some districts represent the bottom quintiles of the population. While wealth is an important focus, the geographic elements of poverty would have been overlooked without disaggregation. Understanding the correlates of poverty will be an important element in reducing it. This analysis implies that in many countries, reducing inequality in health will require tailoring policies to specific geographic areas. For this reason geographic identifiers should be added to all surveys, including MICS and DHS, to allow countries to georeference survey information.

Importantly, different health indicators yielded different patterns of inequity. For example, AIDS knowledge may be high and somewhat equitably distributed, but delivery by skilled birth attendant and U5MR within the same country may be grossly inequitable (as in Cambodia). Inferences about the nature or extent of inequities in health cannot be drawn from a single indicator. Nor can we assume that groups disadvantaged in one indicator are necessarily disadvantaged in another. Reliance on single indicators alone—and certainly national level averages—would lead to limited, misguided recommendations for policy.

The *Guide* stresses the importance of testing for statistical significance and provides an overview of the different types of tests available. Once statistical significance is ascertained, an assessment of the structure of health inequity is also important (see Chapter 6). The actual pattern of inequity may also have policy implications. In one scenario, the indicator may be rather egalitarian across population groups with a marked advantage for just one small subgroup of the population, (as is typical of countries with high or very high mortality for the majority of the population). In these countries, only the very wealthy, often less than 20 percent of the population, can afford and gain access to health services. The primary health care system is practically non-existent for most of the population. Alternatively, the distribution of health inequity may be relatively bimodal—with a high and a low subgroup and all other groups in a relatively similar situation. In this case, differential policies (or redistribution) may be required for one or two regions (or subgroups). Finally, a more complex and heterogeneous situation—such as a stepwise gradient--may indicate the need for other more complex policy combinations.

Chapter 11 addresses the issue of monitoring trends over time. In many ways the evolution of 'health gaps' over time is our main concern. Despite the fact the surveys that formed the bulk of this analysis were only compiled for a single time point for this analysis, future surveys will make these very temporal comparisons feasible. While achieving a perfectly equitable health system may be a long term goal, it not possible, but we argue desirable, that the MDGs and other key development goals be met in an equitable fashion. By 2015, it is imperative that key health indicators like measles immunization, contraceptive use, U5MR and access to skilled birth attendants improve, not only on average, but for all population groups including the most marginalized. This *Guide* presents some initial technical input on how to create the baselines from which such progress toward equity will be measured.

Limitations

The analysis presented herein is in no way meant to substitute for in-depth country-level work on researching inequities in health using local data sources, including vital registration, clinic and census data where it is available. As mentioned above the quantitative results presented here will not be able to fully elucidate the complexities of health systems and health-related behaviors. More in-depth and multivariate quantitative analysis would assist in the clarification of causal pathways that lead certain groups to be disadvantaged relative to others. Qualitative studies will play an important complementary role in explaining patterns and setting priorities. Allowing those who are marginalized and excluded to have a voice in prioritizing their main health concerns is important. Behavioral research, quality of care, levels of trust and cultural norms are additional dimensions of a health system which should be reflected in a holistic equity analysis.

Certain population groups may be completely overlooked in analysis that relies about either population-based surveys or, in some cases, even vital registration. For example, refugee populations fall outside the purview of these data sources, and yet, in many countries represent sizable populations with poor health outcomes and inadequate access to health care. Other groups like orphans and ethnic or linguistic minorities are often not well represented in population-based surveys. Surveys must overcome long-standing shortcomings in the sampling frame such that vulnerable populations are fully assessed. The efforts of MICS to include orphans and of DHS to include family structure and unmarried women are important to capturing all population groups.

Furthermore, measures of the health system are wholly inadequate. This *Guide* uses measures of access to health services that comes from a survey module in which questions about access to family planning and delivery services, and children's health, are asked, and is thus an imperfect indicator. Evaluations of the health system will most likely need to come from a variety of data sources. In an evaluation of constraints to expanding access to health interventions, Ransom et al (2003) use several indicators to estimate health services delivery and health systems management including the following: DPT3 coverage, nurses per 100,000 population, access to health services (a UNICEF measure) and percent case detection of new smear positive cases to assess management. Population based surveys may be able to measure certain aspects of the health system such as household or individuals' distance from the health service, expenditures on health services or judgments about the health system, but assessments of the level of the clinical preparedness of the worker or facility will obviously need to be gathered elsewhere. Task Force 4 of the UN Millennium Project called for the inclusion of 'access to emergency obstetric care (EmOC)' as a necessary complement to the maternal mortality indicator. DHS and MICS do not measure access to EmOC, yet the distribution of this element of the health system is critical to reducing maternal mortality.

In addition, the issue of causality must be addressed here as well. Though certain stratifiers may work well for highlighting health disparities—the direction of causality is not necessarily readily determined. Poverty, for example may cause ill-health but ill-health may be just as likely to cause poverty. Furthermore, though ethnicity may stratify health, we cannot attribute causality solely to ethnicity as another factor such as income or education (or both) may underlie the disparities. The simultaneous stratification allows more nuanced analysis of dual forms of advantage but still conclusions about causality must be made carefully.

In sum, health exclusion results from multiple and overlapping forms of social exclusion, in addition to differences in health systems. The full array of underlying social determinants of health must be addressed in both health research and development policy. And rather than a patchwork of 'pro-poor' interventions and ad hoc targeted programs, universal health systems dedicated to the inclusion of all population groups are needed to build more efficient, equitable and healthier societies.

Chapter 13
Lessons for Implementation and Equity-Sensitive MDGs

The examples presented in this *Guide* confirm that the current focus on pro-poor health policies is an oversimplification that omits other core sources of health inequities (Balk et al 2006). Stratification by wealth, ethnicity, educational level of the mother, sex, region and urban/rural residence yielded statistically significant disparities across a wide range of health indicators in six countries. In many cases, the ethnic, educational and regional variations were more pronounced than the disparities by wealth level. These basic findings imply that dividing the population by wealth quintile alone may be too crude a means of reaching marginalized groups.

Such findings have direct implications for monitoring of the MDGs. Tracking the MDGs in an equity-sensitive manner is both necessary and feasible even with current and limited data. As a start, an (in)equity baseline can be established even in data-poor countries using population-based surveys such as DHS and MICS. Countries can begin with a clear health (in)equity baseline based on the MDGs but tailored to their unique socio-cultural dynamics. Once the (in)equity baseline has been established, the more challenging work begins– linking specific policies to disparity reduction. The MDG targets can be reframed at the country level to prioritize marginalized groups and to reflect the health priorities of the country (and of the marginalized groups). Equitable progress toward the MDG targets implies that the health outcomes of the disadvantaged will improve at the same or faster rates as the better-off groups (Freedman et al 2005, Ngom et al 2003).

There are many moving pieces that must be aligned in order to reach such a goal. This *Guide* outlines the basic technical steps necessary to begin monitoring key child and reproductive health indicators associated with the MDGs. Ideally, country teams and researchers will be able to utilize the *Guide*, to expand upon it and improve it. These techniques may be applied to other MDGs including those related to AIDS, TB and malaria as well as education. A future edition of this *Guide* might reflect empirical results through case studies based upon new efforts to make the MDGs equity-sensitive.

The technical notes detailed herein may well be tempered by the realities of policymaking and negotiating with donors. But it is important that the principles of equity be upheld and be buttressed by sound methods for monitoring progress. This *Guide* is issued in the spirit of ensuring that marginalized groups are buoyed in pursuit of the MDGs.

APPENDIX A

Definitions of indicators, based on MDGs				
ISSUE	**CLOSEST RELATED MDG**			**INDICATOR DEFINITION**
	GOAL	**TARGET**	**INDICATOR**	
Child Mortality	4	5	13	Mortality rates per 1,000 live births ▪ Under age 5 years ▪ Child (under age 5 years; above age 1 year) ▪ Infant (under age 1 year) ▪ Neonatal (under age 28 days)
Underweight	1	2	4	Proportion of children under age 5 years who are moderately or severely underweight ▪ Moderately or severely underweight: below 2 standard deviations below median weight for age of reference population; ▪ Severely underweight: below 3 standard deviations
Knowledge of AIDS	6	18	19b.	Percentage of population aged 15–24 years with comprehensive correct knowledge of HIV/AIDS (UNICEF/WHO) ▪ Women aged 15-24 who know that a healthy-looking person can transmit HIV, percent (UNICEF/UNAIDS/WHO) ▪ Women aged 15–24 who know that a person can protect herself from HIV infection by consistent condom use, percent (UNICEF/UNAIDS/WHO)
Contraceptive Prevalence	6	18	19c.	Contraceptive prevalence rate, proportion of currently married women aged 15–49 using contraception (UN Population Division) ▪ Any method ▪ Condom ▪ Modern method (DHS definition of modern method: pill, IUD, injections, condom, female sterilization, male sterilization, implants, lactational amenorrhea, foam or jelly, emergency contraception; does not include withdrawal, periodic abstinence or folk methods; MICS definition similar)
Measles	4	5	15	Proportion of 1 year-old children immunized against measles (UNICEF/WHO)
DPT Vaccine				Proportion of 1 year-old children immunized against DPT (3 doses)
Delivery Care	5	6	17	Proportion of births attended by skilled health personnel (UNICEF/WHO). Refers exclusively to people with midwifery skills (for example, doctors, midwives, nurses) who have been trained to proficiency in the skills necessary to manage normal deliveries and diagnose or refer obstetric complications
Age at First Intercourse				Average age at first intercourse
Age at First Marriage				Average age at first union
Health System Use				Proportion of women who visited a health facility in the past 12 months for self or child

APPENDIX B
Further Explanation of Pivot Tables

While pivot tables are time-intensive to create, they provide excellent means to isolate certain data for comparison across time, country, stratifier or indicator. A large amount of data can be included in a pivot table, but much can be hidden if it is not relevant at a given time. For this analysis, SPSS was used to make cross-tabulations of the data and then this information was copied into Microsoft Excel. Tables can be pivoted within SPSS, but to achieve full flexibility, more complex programming is required. This Appendix provides a brief overview of how to create and use pivot tables.

In Excel, the cross-tabulation output (the input to the pivot table) should be set up in rows so that each row corresponds to a unique country, stratum, and indicator. Multiple countries, stratifiers and indicators can be included in the same list as multiple rows and then reordered to optimize comparisons in a pivot table. The example below shows a selection of the input for a singly stratified analysis of Cambodian facility use. The number of cases is useful for suppressing values based on a small number of cases, but is not otherwise used.

Table B1: Example of Excel Preparation for a Pivot Table

INDICATOR	COUNTRY	CLASS	LEVEL	CASES	PERCENTAGE
Visited Health Facility in Last Year	Cambodia	National Average	n/a	15344	11.2
Visited Health Facility in Last Year	Cambodia	Education	None	4848	9.4
Visited Health Facility in Last Year	Cambodia	Education	Primary	8178	10.8
Visited Health Facility in Last Year	Cambodia	Education	Secondary or more	2318	15.8

Once the data is entered into Excel in this format, choose ‘PivotTable and PivotChart Report’ from the ‘Data’ Menu. Select the whole table in the resulting dialogue box, and place the new Pivot Table on a new worksheet. In order to display the highlighted values in the first column of Table B2, drag the Class and Level fields from the PivotTable Field List to the ‘Drop Row Fields Here’ area, Indicator to the ‘Drop Column Fields Here’ area, Country to the ‘Drop Page Fields Here’ area, and Percentage to the ‘Drop Data Items Here’ area.

Table B2: Example of a Pivot Table in Excel

CAMBODIA								
		INDICATOR						
Stratification Class	Stratification Level	ACCESS TO HEALTH CARE FACILITY (VISITED IN LAST YEAR)	AGE AT FIRST INTERCOURSE	AGE AT FIRST MARRIAGE	AIDS KNOWLEDGE - A HEALTHY PERSON HAVE AIDS	AIDS KNOWLEDGE - USE CONDOMS DURING SEX	CPR - ANY METHOD	CPR - CONDOM
National Average		**11**	**19.3**	**19.4**	**87**	**82**	**14**	**1**
Education	**P-value*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	None	9	19.1	19.3	87	72	13	0
	Primary	11	19.2	19.3	86	83	14	1
	Secondary or more	16	20.0	20.1	92	91	18	1

The drop-down menus within each column of the pivot table allow users to show all the data or limited selections in order to compare numbers more easily side-by-side, so that large discrepancies or irregularities can stand out. By isolating certain data, important trends will not get lost in large tables.

Simultaneously stratified data can be pivoted analogously. Additional stratification class and stratification level fields must be populated in the input table and then displayed in the output table. The use of color-coding can be especially helpful in identifying stratification classes here.

The pivot table function is also available in programs such as SPSS. If you prepared the data in SPSS, it can be simpler to leave the data in that program, as shown in the table below. However, the options are fewer in SPSS and less can be done in terms of presentation.

Table B3: Example of SPSS Output for Pivot Tables

Report

Mean

Education	Visited health facil. last 12m	Age at first intercourse	Age at first marriage	Can a healthy person have AIDS	AIDS: use condoms during sex	Any Method of Birth Control	Birth Control - Condom
None	.09	19.10	19.28	.87	.72	.13	.00
Primary	.11	19.16	19.29	.86	.83	.14	.01
Secondary or more	.16	19.99	20.12	.92	.91	.18	.01
Total	.11	19.26	19.40	.87	.82	.14	.01

Appendix C: Methods of calculating child mortality rates from survey data

Depending on the nature of the survey data used, there may be some flexibility in the choice of method for calculating infant and child mortality for population subgroups. Broadly speaking, two methods are in general use. Direct methods require information on the survival status of each birth up to the age in question, the age at death of children who died, and the age of children still alive. This information can be difficult to obtain, but if done so accurately, will provide good estimates.

The most efficient indirect methods require only three pieces of information from each mother: her age, the number of children ever born, and the number who have died before a specific age (Brass 1964; Sullivan 1972; Feeney 1976; United Nations 1983). Additional information can be used in some hybrid approaches (Preston and Palloni 1978).

DHS surveys provide enough information to use either method, but they report based on direct methods (Rutstein and Rojas 2003). DHS distributes programs for SPSS and SAS that calculate mortality and fertility rates for a given population. However, they require a substantial amount of additional programming to produce estimates for multiple non-overlapping population groups at the same time. The SPSS program also requires the Tables add-on module.

MICS surveys only have enough information to produce estimates based on indirect methods. UNICEF distributes a free DOS program called QFIVE, (UN Population Division 1990) that requires simple cross-tabulations of all births and deaths by mothers' ages as input. Like the DHS program, it is only readily able to calculate rates for one population group at a time.

Mortality rates, especially those of neonates, are more sensitive to low sample sizes, as they are generally lower than other measured rates. Because of this, DHS does not report rates based on less than 250 live births, and urges caution in the interpretation of results based on less than 500. These thresholds are ten times those used for most other statistics.

REFERENCES

Anand, S. et al. "Measuring Disparities in Health: Methods and Indicators." In Evans, T., Whitehead, M, et al. (2001). *Challenging Inequities in Health: From Ethics to Action*. New York: Oxford University Press.

Balk D. (1997). Defying gender norms in rural Bangladesh: a social demographic analysis. *Population Studies*. **51**(2): 153–172.

Balk D, Wirth M, Delamonica E, Storeygard A, Sacks E, Minujin A. (2006) Multidimensional disparities in maternal and child health: Measuring a baseline, monitoring the future. In: Minujin A, Delamonica E, Komarecki, M. eds. *Poverty and Children. Policies to break the vicious cycle.* New York: New School University Press, forthcoming, 2006.

Bambas Nolen, L., Braveman, P., Dachs, J.N.W., Delgado, I., Gakidou, E., Moser, K. et al. (2004). Strengthening health information systems to address health equity challenges. *Bulletin of the World Health Organization*, **83**:597-603.

Brass W. (1964). Uses of census and survey data for the estimation of vital rates. Paper prepared for the African Seminar on Vital Statistics, Addis Ababa, 14-19 December. E/CN.14/CAS.4/VS/7.

Braveman, P. (1998). *Monitoring Inequities in Health: A Policy-Oriented Approach in Low- and Middle-Income Countries*. WHO/CHS/HSS/98.1. Geneva: World Health Organization. http://whqlibdoc.who.int/hq/1998/WHO_CHS_HSS_98.1.pdf

Braveman P. and Gruskin S. Poverty, equity, human rights and health. (2003). *Bulletin of the World Health Organization*, **81**(7):539-545.

Brockerhoff, M. and P. Hewitt. (2000). Inequality of child mortality among ethnic groups in sub-Saharan Africa. *Bulletin of the World Health Organization.* **78**(1):30-41.

DHS Dimensions. (2000). New Directions: DHS Surveys incorporate Geographic Data. DHS Vol 2(1) Spring Issue.

Elledge, M., Bloom, E. et al. (2001). *Health and Education Needs of Ethnic Minorities in the Greater Mekong Subregion*. Manila, Philippines: Asian Development Bank.

Feeney G. (1980). "Estimating infant mortality trends from child survivorship data." *Population Studies*,**34**(1):109-128.

Filmer, D. and Pritchett L. (2001). Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. *Demography*. **38**(1):115-32.

Freedman L, Waldman R, de Pinho H, Wirth, M, Chowdhury AMR, Rosenfield, A. (2005). *Who's got the power? Transforming health systems for women and children.* UN Millennium Project Task Force on Child Health and Maternal Health. London, Earthscan.

Freedman, L., Wirth, M. et al. (2004). *Interim Report. Task Force 4: Child health and maternal health*. New York: UN Millennium Project.

Gwatkin, D.R. (2002). *Who would gain most from efforts to reach the Millennium Development Goals for health?: An inquiry into the possibility of progress that fails to reach the poor*. HNP Discussion Paper. Washington, DC, The World Bank.

Gwatkin, D., Rutstein, S. et al. (2003). *Initial country-level information about socio-economic differences in health, nutrition, and population.* Washington, DC, The World Bank.

Gwatkin, D.R., Rutstein, S., Johnson, K., Suliman, E.A., Wagstaff, A. (2003). *Initial country-level information about socio-economic differences in health, nutrition, and population.* Washington, DC, The World Bank.

Komarecki, M. (2003). Report to UNICEF: Promoting human rights and social policies for children and women: the role of Multiple Indicator Cluster Survey (MICS). The New School University, New York.

McCloskey, D. and Ziliak, T. (1996) 'The Standard Error of Regression', *Journal of Economic Literature*, March.

Minujin, A. and Delamonica, E. (2003a). Mind the Gap! Widening Child Mortality Disparities. *Journal of Human Development*. **4**(3): 397-418.

Minujin, A. and Delamonica, E. (2003b) *Equality matters for a world fit for children: Lessons from the 90s*. UNICEF Staff Working Papers, Division of Policy and Planning Series (Draft, Number DPP-03), New York, UNICEF.

Moyo, S. (2004). *Socio-Economic Dominance of Ethnic and Racial Groups – The African Experience*. Human Development Report Office (Occasional Paper, Background Paper for HDR 2004). United Nations Development Programme, 2004/8.

Preston, S.H. and Palloni, A. (1978) "Fine-tuning Brass-type mortality estimates with data on ages of surviving children." *Population Bulletin of the United Nations*, No. 10-1977, 72-87.

Rutstein, S.O. and Rojas, G. (2003). Guide to DHS Statistics. Calverton, Maryland: ORC Macro.

Sachs, J. et al. (2004). Millennium Development Goals Needs Assessments: Country Case Studies of Bangladesh, Cambodia, Ghana, Tanzania and Uganda. Working Paper. New York: UN Millennium Project.

Sahn, D.E. and Stifel, D.C., (2000). "Poverty Comparisons Over Time and Across Countries in Africa," *World Development*, Elsevier, vol. **28**(12), pages 2123-2155.

Sullivan, J.M. (1972) "Models for the estimation of the probability of dying between birth and exact ages of early childhood." *Population Studies*. **26**(1): 79-97.

United Nations. (1983). Manual X: Indirect Techniques for Demographic Estimation. Department of International Economic and Social Affairs. *Population Studies* No. 81. New York.

UN Millennium Project. (2005). *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals. Overview*.

United Nations Development Programme (UNDP). (2003). *Human Development Report: Millennium Development Goals: A Compact among Nations to End Poverty*. *2003*. New York: UNDP.

UNDP (United Nations Development Programme). (2001). *Human Development Report 2001: Making Technology Work for Human Development*. New York, Oxford University Press.[http://hdr.undp.org/reports/global/2001/en/].

United Nations Population Division (1990). Step-by-step guide to the estimation of child mortality/ New York: United Nations.

Vandemoortele, J. (2000). *Absorbing social shocks, protecting children and reducing poverty: The role of basic social services*. UNICEF staff working papers, Evaluation, Policy and Planning Series no. 00-001. New York: UNICEF.

Vega, J., and Irwin, A. Tackling health inequalities: new approaches in public policy. *Bulletin of the World Health Organization*, 2004, **82**(7): 482.

Wirth, M., Balk, D., Storeygard, A., Sacks, E., Delamonica, E., Minujin, A. (2004). Setting the stage for equity-sensitive monitoring of the health MDGs. Background paper commissioned by the UN Millennium Project Task Force on Child Health and Maternal Health. New York.

Wirth, M., Balk, D., Delamonica, E., Storeygard, A., Sacks, E., Minujin, A. (2006). Setting the Stage for Equity-Sensitive Monitoring of the Health MDGs. *Bulletin of the World Health Organization*. Forthcoming.

World Health Organization. (2005). Towards a Conceptual Framework for Analysis and Action on the Social Determinants of Health. Discussion paper for the Commission on Social Determinants of Health – CSDH. DRAFT 5 May 2005. WHO Health Equity Team, Office of the Assistant Director-General, Evidence and Information for Policy Cluster. http://ftp.who.int/eip/commision/Cairo/Meeting/CSDH%20Doc%202%20-%20Conceptual%20framework.pdf. Accessed October 16, 2005.

DATA REFERENCES

Central Statistical Authority [Ethiopia] and ORC Macro. (2001). *Ethiopia Demographic and Health Survey 2000*. Addis Ababa, Ethiopia and Calverton, Maryland, USA: Central Statistical Authority and ORC Macro.

Centro de Estudios Sociales y Demográficos (CESDEM) and MEASURE DHS+/ORC Macro. (2003). *República Dominicana Encuesta de Demografía y Salud 2002*. [Dominican Republic Demographic and Health Survey 2002] Calverton, Maryland, USA, Centro de Estudios Sociales y Demográficos, CESDEM and MEASURE DHS+/ORC Macro. In Spanish.

Ghana Statistical Service (GSS) and Macro International Inc. (MI). (1999). *Ghana Demographic and Health Survey 1998*. Calverton, Maryland: GSS and MI.

National Council for Population and Development (NCPD), Central Bureau of Statistics (CBS) (Office of the Vice President and Ministry of Planning and National Development) [Kenya], and Macro International Inc. (MI). (1999). *Kenya Demographic and Health Survey 1998*. Calverton, Maryland: NDPD, CBS, and MI.

National Institute of Statistics, Directorate General for Health [Cambodia], and ORC Macro. (2001). *Cambodia Demographic and Health Survey 2000.* Phnom Penh, Cambodia, and Calverton, Maryland, USA: National Institute of Statistics, Directorate General for Health, and ORC Macro.

United Nations Children's Fund (UNICEF). (2000). *Multiple Indicator Cluster Survey (MICS): Tajikistan*. Dushanbe, UNICEF.