Published in final edited form as:

Soc Sci Med. 2014 April ; 106: 143–150. doi:10.1016/j.socscimed.2014.01.045.

CHALLENGES OF HEALTH MEASUREMENT IN STUDIES OF HEALTH DISPARITIES

Sarah A. Burgard^a and Patricia V. Chen^b

^bPatricia V. Chen, University of Michigan Department of Sociology, 500 South State Street, Ann Arbor, Michigan, 48109-1382, U.S.A., pvchen@umich.edu.

Abstract

Health disparities are increasingly studied in and across a growing array of societies. While novel contexts and comparisons are a promising development, this commentary highlights four challenges to finding appropriate and adequate health measures when making comparisons across groups within a society or across distinctive societies. These challenges affect the accuracy with which we characterize the degree of inequality, limiting possibilities for effectively targeting resources to improve health and reduce disparities. First, comparisons may be challenged by different distributions of disease and second, by variation in the availability and quality of vital events and census data often used to measure health. Third, the comparability of self-reported information about specific health conditions may vary across social groups or societies because of diagnosis bias or diagnosis avoidance. Fourth, self-reported overall health measures or measures of specific symptoms may not be comparable across groups if they use different reference groups or interpret questions or concepts differently. We explain specific issues that make up each type of challenge and show how they may lead to underestimates or inflation of estimated health disparities. We also discuss approaches that have been used to address them in prior research, note where further innovation is needed to solve lingering problems, and make recommendations for improving future research. Many of our examples are drawn from South Africa or the United States, societies characterized by substantial socioeconomic inequality across ethnic groups and wide disparities in many health outcomes, but the issues explored throughout apply to a wide variety of contexts and inquiries.

Keywords

Health disparities; health measurement; diagnosis bias; self-rated health; cross-national comparison

© 2014 Elsevier Ltd. All rights reserved.

^a[Corresponding author] University of Michigan Department of Sociology, 500 South State Street, Ann Arbor, Michigan, 48109-1382, U.S.A., burgards@umich.edu, Telephone 734-615-9538. .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

Interest in understanding the nature, extent, and causes of health disparities has expanded dramatically. Historically, many studies focused on North America and Europe, but increasingly, health disparities are studied within and across a wider array of societies. The growing diversity of contexts studied highlights the difficulty of finding appropriate and adequate measures of health when comparing across groups or societies. These challenges affect the accuracy with which we characterize the degree of inequality, limiting possibilities for effectively targeting resources to improve health and reduce disparities. This commentary highlights four common challenges: different distributions of disease, variation in the availability and quality of vital events and census data, variation in the comparability of self-reported information about specific health conditions due to diagnosis bias or avoidance, and challenges to comparability of self-reported overall health or symptoms measures.

Other challenges of measurement and comparability have been noted in health research, such as difficulties measuring race/ethnicity or socioeconomic status, or the mechanisms underlying disparities, such as discrimination or acculturation (Murray & Frenk, 2008; Stewart & Nápoles-Springer, 2003). We focus instead on the potential pitfalls of using commonly-available measures of health when comparing diverse groups. Earlier studies pursuing a variety of research aims with different comparison populations have noted some aspects of the challenges we discuss (see for example Bardage et al., 2005; Bramley et al., 2005; Lima-Costa et al., 2012; Murray & Frenk, 2008). Our contribution is a more focused commentary on the measurement of health in studies of health disparities, with clear recommendations for improving future research. We explain specific issues that make up each type of challenge and show how they may lead to underestimates or inflation of estimated health disparities. We also discuss approaches that have been used to address them in prior research and note where further innovation is needed to solve lingering problems.

We do not attempt an exhaustive literature review, but highlight these common challenges with clarifying exemplars, many drawn from South Africa or the United States, nations with substantial socioeconomic inequality across ethnic groups and wide disparities on many health outcomes. These societies clearly exemplify challenges faced by researchers who want to consider within-nation health disparities generated by politicized and/or racialized material resource divides. South Africa in particular is a context where socioeconomic and epidemiologic characteristics of wealthy and low-income societies coexist, making it particularly illustrative of challenges to both within- and across-society comparisons. South Africa and the U.S. also are cases that challenge social scientists to be mindful of the potential limitations of available measures of health that have been strongly influenced by systems of power unique to each society. Historically- and socially-embedded systems of inequality have shaped health policy and institutionalized practices that underlie both the disparities in health outcomes under study and the availability and quality of information about the health of differently-situated groups (Krieger, 2011). While these societies are particularly apt for illustration, we also use examples from other societies and comparisons. We draw mainly from sociology, demography and social epidemiology to frame our

commentary, but the issues and examples explored throughout apply to a wide variety of contexts, perspectives, and inquiries.

CHALLENGE ONE: DIFFERENCES IN DISEASE DISTRIBUTIONS ACROSS GROUPS

Description of the problem and implications

Differences in the distribution of disease across societies and across groups within societies can make it difficult to find a common health metric on which to compare the affluent and the poor. Traditionally, diseases associated with poverty have included communicable diseases such as tuberculosis, malaria, and infectious diseases of childhood (DFID, 2000), and morbidity related to maternal, perinatal and nutritional problems. By contrast, chronic diseases, poor mental health, and harmful health-related behaviors dominate the disease profile of wealthier, more industrialized societies, complicating cross-national comparisons of health. Profiles of disease are in transition for any society; with macroeconomic development and diffusion of better infrastructure and medical care, they experience “epidemiologic transition.” In the classic rendering, a profile of disease dominated by communicable diseases and conditions associated with malnutrition or lack of basic infrastructure transitions to a profile dominated by chronic disease (Omran, 1971). However, more recent studies show that the disease profiles of lower-income societies are much more complicated than Omran’s account of universal, linear epidemiologic transition would suggest. While infectious disease and other “diseases of poverty” continue to be a burden, chronic diseases increasingly are taking a toll on low-income populations (World Health Organization 2005).

Moreover, epidemiologic transition has not proceeded at the same rate across groups within some societies. The burden of communicable diseases, malnutrition and maternal and perinatal problems may linger for poor groups even while chronic diseases also begin to affect them (Schneider et al., 2009; Steyn & Schneider, 2001). Uneven epidemiologic transition has been labeled “protracted-polarized transition” (Frenk et al., 1989) and characterizes some contemporary middle- and lower-income societies. Some have argued that uneven transition within societies is not random or temporary, but rather may reflect underlying systems of power in a given society and may thus persist if it remains beneficial to more advantaged groups (Krieger, 2011). Complicating the picture further, some societies have experienced “counter transitions” when emergent communicable diseases like HIV/AIDS reverse recent mortality and morbidity gains (Kahn et al., 2007), and potentially widen gaps between advantaged and disadvantaged groups.

Difficulties comparing the health of societies or social groups with different profiles of disease burden could lead to over- or under-estimates of health disparity, depending on the specific health outcome considered. While chronic diseases may affect all groups to some degree, focusing on chronic conditions underestimates the total burden of disease for the poor, who also still cope with diseases of poverty. Moreover, earlier mortality of poorer individuals means that they are less likely to survive long enough to be at risk for some chronic diseases that manifest later in life (Willson et al., 2007), so their burden of chronic

disease could appear artificially low. Conversely, a focus on diseases of poverty, such as diarrheal disease deaths, would show stark disparity between advantaged and disadvantaged groups but likely be unrepresentative of the overall level of health inequality because disparities for other health problems are narrower.

Strategies for addressing the problem

At a minimum, researchers should consider the suitability of a particular health indicator for comparison across groups or societies characterized by very different distributions of disease. As one solution to this challenge of comparability, overall mortality rates have been used to gauge the relative survival across societies or groups within societies, but do not capture considerable nuance in health status among survivors. Alternatively, researchers have compared across multiple indicators of morbidity and mortality, as does a study of disparities in mortality rates, life expectancy, coronary heart disease and diabetes prevalence, self-rated health, activity limitation and sedentariness that shows “...pervasive —albeit not invariable — patterns suggesting incremental income or education gradients for a range of important health indicators among both children and adults” in the U.S. (Braveman et al., 2010: S189). However, in some contexts data may not be available on such a wide range of reliable outcomes and even where they are available, results from a variety of indicators may not neatly or conclusively show consistent disparities (Braveman et al., 2010).

An alternative measure that attempts to summarize the relative health standing of societies with very different profiles of disease is the disability adjusted life year (DALY), (World Health Organization, 2009). Health economists created the DALY measure to quantify the global burden of disease (Murray et al., 2002). DALYs integrate years lost from premature mortality and from living with a disability to arrive at a summary measure of the health status of a population or group (Mont, 2007). This makes it useful for calculating and comparing burdens of disease that arise from a broad array of factors contributing to poor health (e.g., malnutrition, chronic diseases, substance use), and thus addressing the challenges of comparing societies or groups within societies that are at different stages of epidemiologic transition. Nonetheless, the DALY measure has been critiqued for being “blind” to contextual and socioeconomic factors that modify how an individual may experience morbidity, making it unclear whether a given condition should count as “disabling” in every context (Mont, 2007). For example, a health condition limiting the ability to carry heavy objects or work long hours in a standing position may restrict employment and financial well-being in an agriculturally-based economy or for some jobs in industrialized economies, and thus may justifiably represent disability. However, the same condition may not affect employability and access to resources in a more service-based economy or occupation, or where mobility aids are widely available. Others have critiqued DALY measures for being driven mainly by mortality rates, leading to underestimates of the burden of many tropical diseases of poverty that do not have high mortality rates but generate major health burdens including blindness or disability (Conteh et al., 2010). Finally, reliance on the DALY may complicate allocation of health intervention because it averages across specific causes, potentially obscuring specific health targets (Anand & Hanson, 1997; Mont, 2007) or minimizing burdens associated with causes borne

disproportionately by disadvantaged groups. For these and other reasons, DALYs may not always be useful for measuring disparities in health. However, they have been applied widely as one approach to problems of comparability across very different contexts (for discussion of other summary metrics, see Braveman, 2006).

CHALLENGE TWO: VARIATION IN AVAILABILITY/QUALITY OF VITAL EVENTS OR CENSUS DATA

Description of the problem and implications

Data availability and quality are a challenge in studies of health disparities. One of the most basic problems is differential availability and quality of vital events data (birth, death, and marital records) and population census data that are often used to study disparities in overall mortality and causes of death. In low-resourced societies, vital events records or population censuses may not exist, have only recently been initiated, or suffer from low quality. Demographers have characterized two key issues with vital event and census data: coverage errors and content errors. Coverage errors occur when some events or individuals are not enumerated, while content errors reflect missing or incorrect characteristics on a record that is obtained (Siegel & Swanson, 2004).

Coverage and content of vital events and censuses may be flawed for entire populations in societies lacking data collection resources, or they may vary across groups within societies in ways that reflect power gradients and political decisions. For example, Steyn and Schneider (2001) note that during the institutionalized racism of the Apartheid era in South Africa, deaths of Blacks in rural areas often went unreported. For reported deaths, cause was often misclassified as “ill-defined” due to a lack of resources for more careful investigation. Coverage and content of vital events was better for South African whites and urban residents, reflecting their privileged status, including political prioritization and protected residence in urban areas with better public health infrastructure. Even in the relatively well-resourced U.S., studying historical disparities between African Americans and whites is challenging. Southern U.S. states, where many African Americans resided, were among the last to implement mandatory birth registration (limiting information needed to calculate some vital rates). Also, because the African American population tended to have more mobile household structures, it was more difficult to obtain accurate annual estimates of their place of residence (a limitation affecting the denominators of calculated rates of mortality distribution in 1985) (Preston et al., 1996). If vital statistics are more complete and accurate for advantaged groups or societies, deaths of disadvantaged groups may be undercounted, resulting in underestimates of mortality disparities. Poorer quality cause-of-death data for some groups could also affect estimates of disparities for specific causes.

Strategies for addressing the problem

The optimal solution for problems of data coverage and content is to improve the quality of vital registration and census data collection systems, but such developments involve enormous investments of time and resources and are typically beyond the control of researchers. While political attention should be directed to the importance of addressing these limitations that often underestimate the health burden of disadvantaged groups

(Marmot et al., 2008), researchers have developed approaches to data limitations when more adequate data are unavailable. One approach takes advantage of existing population censuses in locations lacking well-functioning vital events registries, applying modeling techniques such as life tables or stable population models to these census data to obtain estimates of age-specific mortality (Siegel & Swanson, 2004). In contexts without reliable information on age distributions or other key inputs to life table approaches, model life tables for global regions or a similar society are used.

Another approach supplements existing vital events or census data with information from sample surveys (cross-sectional or longitudinal), sample vital registrations, or demographic surveillance sites. Population-based survey sampling can provide accurate estimates with much smaller samples and at greatly reduced cost compared to population censuses, and can be more frequently conducted and tailored to emergent measurement needs. Using modeling methods, researchers can patch together survey data and vital events or census data to obtain mortality rates and even rates for specific causes of death (Siegel & Swanson, 2004). Of course, survey data are themselves not immune from error; potential problems with survey data are discussed below.

An alternative to using survey data to supplement vital events or census data is reliance on surveillance sites that carefully track the vital events of a relatively small community frequently and over many years. The long-running Agincourt study site in the rural northeast of South Africa is an example, covering about twenty villages with 70,000 residents. Interviews conducted in 1992 and 2003 used different survey modules, asked about additions and departures from households, and investigated deaths with detailed verbal autopsies (Kahn et al., 2007). Such health and demographic surveillance sites (HDSS) have been established in many middle and low-income societies around the world (Sankoh & Byass, 2012). While these communities may not represent the general population and the data may be of limited usefulness for comparing across social groups or societies, surveillance sites provide otherwise unavailable information for some health measures.

CHALLENGE THREE: PROBLEMS WITH SELF-REPORTS OF HEALTH CONDITIONS

Description of the problem and implications

In addition to vital events, census, and surveillance data, surveys also provide health data across a range of societies. Surveys are generally more flexible than vital events registries or censuses, adaptable to measuring changing trends in health or social disadvantage, and can be used to collect a wider breadth of information on a more representative sample of the population than surveillance sites. However, surveys often rely heavily on self-reports about health and other indicators. If the goal is measuring a physiologically-universal health condition across groups that differ on resources or cultural practices, self-reported information is vulnerable to at least two potential issues that could reduce comparability: diagnosis bias and diagnosis avoidance.

Diagnosis bias stems from unequal access to medical providers or health information and potentially distorts the observed prevalence of health problems among survey respondents.

For example, to assess the prevalence of hypertension, surveys commonly ask questions like: “Has a doctor ever told you that you have hypertension or high blood pressure?” Incorrect estimates of disparities could emerge if respondent groups have differential access to medical care and diagnoses. Advantaged groups with better access to quality health care are probably more aware of these health conditions and are socially-situated in ways that makes it easier to answer survey questions designed by researchers who also tend to be socially-advantaged. By contrast, substantial portions of the population in some societies lack even basic access to medical care, such as disadvantaged women in poorer countries including India, Nepal, Bangladesh, and parts of the Middle East (Ojanuga & Gilbert, 1992). Access issues could lead to low awareness of some of health conditions and/or inability to literally report that “a doctor told” the respondent about their condition. Even among those with access to health care, limited health literacy among educationally-disadvantaged groups could create barriers to adequate knowledge about health problems (Williams et al., 1998) or the ability to report them in response to survey questions. Differences in access to high quality medical care or health literacy could lead to underestimates of specific health conditions among the disadvantaged.

Differential quality of health care could also result in diagnosis biases. Hospitals serving less-advantaged populations tend to face a greater patient loads, substandard infrastructure, and limited staffing, leading to greater risk of less-thorough examinations and missed opportunities for screenings. Differential treatment of some social groups also could lead to diagnosis bias; U.S. studies using data from the 1980s through 2000 found that racial minority groups do not always receive the most technologically-advanced treatments for their conditions (Epstein, 2004; Felix-Aaron et al., 2006), and differential treatment may extend to diagnostic technologies. Global inequalities in access to care and diagnoses make diagnosis bias an issue for cross-national comparisons as well as for comparisons across groups within societies.

In addition to the problem of diagnosis bias, individuals’ avoidance of diagnoses or medical care may compromise the comparability of self-reports about health conditions. Avoidance of diagnoses may be inadvertent or intentional. Inadvertent avoidance could occur if individuals who feel healthy choose not to allocate scarce resources to see a doctor for routine care, potentially missing conditions that are asymptomatic or not troublesome, leading to an underestimate of disease among less advantaged groups. Those facing fewer budgetary restrictions may be more likely to obtain preventive care even if they feel healthy, making them more likely to be diagnosed and biasing estimates of disparity. For example, a study on self-reported hypertension using Health Survey for England data from 1998 and 2003 found that men with lower incomes and those living outside of the central-urban areas were more likely to be unaware that they had hypertension (Johnston et al., 2009). Inadvertent avoidance also could arise due to cultural differences in help-seeking. For example, Asian Americans are less likely to use mental health services than other ethnic groups in the U.S., and some have argued that reluctance to seek services is a better explanation than lack of need (Zhang et al., 1998).

It is important to note that diagnosis bias may not only affect socially-disadvantaged groups. For example, a 1991 study conducted in the Netherlands compared self-reports with

objectively-collected diagnoses and found that misreporting did not always occur most among the less-advantaged. Rather, some conditions, such as chronic lung disease, were underreported more by those with higher education (Mackenbach et al., 1996). A study in New York from 1997-1999 found more accurate self-reports from disadvantaged groups due to health interventions and targeted screening directed towards disadvantaged communities and health conditions known to disproportionately affect them, such as diabetes (Hosler et al., 2005).

In other cases, diagnosis avoidance is intentional. Individuals may avoid testing for stigmatized conditions, either for social reasons or because a positive diagnosis would affect access to health insurance, death benefits, or other resources. One such case is HIV/AIDS (Kellerman et al., 2002; Van Dyk & Van Dyk, 2003); despite easy diagnosis and considerable resource allocation to encourage testing, HIV/AIDS remains underdiagnosed. The true impact of HIV/AIDS on population health patterns is difficult to discern because of the stigma associated with the disease – some individuals do not learn their status, others do not disclose it, and those who certify deaths sometimes avoid reporting it on death certificates. For example, some estimates suggest that about 90% of HIV/AIDS deaths from 1996-2006 in South Africa were attributed to other causes and thus were massively undercounted (Birnbaum et al., 2011).

At the same time, targeted surveillance of more stigmatized conditions such as HIV/AIDS in South Africa (Kranzer et al., 2011; Tabana et al., 2012) or tuberculosis in Zambia (Ayles et al., 2009) sometimes provides more accurate data for these health conditions among disadvantaged groups. Much of the HIV/AIDS screening in lower-income nations, for example, has been conducted in public antenatal and delivery care facilities that serve childbearing-aged women, or small rural surveillance sites. Such surveillance data may also be inadequate for studying health disparities in stigmatized conditions — not because of poor quality, but because they do not represent the entire population. Undercounting among advantaged groups could lead to overestimates of disparity. For example, knowledge of HIV status was lower for more advantaged women in one 2005 U.S. study; advantaged women were not likely to be offered rapid HIV testing when they were in labor, lowering the opportunity to prevent HIV transmission to the infant (Gaur et al., 2010).

Strategies for addressing the problem

There are several possible solutions to the problems of diagnosis bias and diagnosis avoidance. Some researchers have turned to collecting more “objective” measurements; for example, population-based survey projects have begun collecting biomarkers of specific health problems or risk factors. For example, the 2006 wave of the longitudinal Health and Retirement Study of U.S. adults included saliva collection for DNA extraction and blood spot collection for cholesterol, hemoglobin A1C, CRP and cystatin C analysis (Crimmins et al., 2008). Other biomarker collections have been focused on communicable and stigmatized conditions, chiefly HIV/AIDS in less wealthy nations (Ayles et al., 2009; Kranzer et al., 2011; Tabana et al., 2012). However, biomarkers or other “objective” indicators of specific conditions may not be useful for capturing some critical health conditions, such as mental health problems.

Moreover, while they can address some of the limitations of self-reported measures, biomarkers remain too expensive or challenging to collect for many projects and contexts, or for a broader sample of the general population. With the exception of research and intervention projects aimed at controlling serious communicable diseases, collection of biomarker data has thus been even less plausible in lower-resourced settings (Boerma & Stansfield, 2007). To work around these issues, researchers focused on using more “objective” health indicators have turned to other data sources. Some studies conducted in the U.S. have linked surveys with administrative data to capture measures of health care received or diagnoses for particular conditions plausibly indicated by specific medication prescriptions (Blumberg et al., 2012; Dwyer, 2009; Hajjar & Kotchen, 2003). Others have obtained diagnostic reports or medical records, hospital or Medicare expenditures data, or other administrative data to study disparities. While variation in access to care or diagnosis avoidance may make administrative data incomplete, such data can in some cases be used to measure inequality in targeted health outcomes, particularly in societies where most residents have some access to medical care.

Alternatively, simulations have been conducted to improve prevalence estimates weakened by diagnosis bias or avoidance. For example, one study used 1996-2006 data from the United Nations World Population Prospects to compare age and sex-specific mortality rates in South Africa with rates in similar societies not burdened by major HIV epidemics, illustrating the possible influence of HIV/AIDS on estimates and attributions of morbidity and mortality (Birnbaum et al., 2011). Researchers have also used verbal autopsies collected between 1992 and 2005 at South African surveillance sites and physician coding or statistical algorithms to adjust estimates of deaths from other causes that should have been attributed to HIV/AIDS (Byass et al., 2011).

CHALLENGE FOUR: REFERENCE GROUPS AND COMPARABILITY OF SELF-REPORTED HEALTH AND SYMPTOMS ACROSS GROUPS

Description of the problem and implications

Because of the ease and affordability of collection, its ability to predict mortality (DeSalvo et al., 2005; Idler & Benyamini, 1997), and given the known difficulties in collecting measures of specific health conditions, self-rated health is a popular health measure. Surveys ask respondents: “In general, how would you rate your health - excellent, very good, good, fair, or poor?” or a similar item to gauge overall health. They may also ask respondents to report their own physical or mental health symptoms. However, research using self-reported measures of overall health or symptoms to study highly-stratified societies has yielded some unexpected findings. For example, one study showed that states in India with the highest literacy rates and longevity had the highest morbidity reporting, while those with low life expectancy and poor health and educational infrastructure had the lowest rates of self-reported morbidity (Sen, 2002). Another study found that a lower percentage of Aboriginal people in Australia reported fair or poor self-reported health status compared to the general population, despite their socially-disadvantaged status (Murray et al., 2002).

Such counter-intuitive findings lead us to our fourth point: the influence of reference groups and cultural differences on comparability of self-reported measures of general health or symptoms. The unexpected associations mentioned above could arise because different social groups may use varying reference points when they judge their relative well-being. Different interpretations of questions about overall health or symptoms, different salience of these across groups, or differences in understanding of such survey questions may influence disparities estimates.

Individuals are likely to consider the health of the people around them when making judgments about their own health in reporting their overall well-being or symptoms. Less-advantaged individuals' levels of health expectation may be lower than those of more-advantaged individuals because health problems are more common or even expected in poorer communities (Dahlgren & Whitehead, 2007). A study comparing U.S. respondents' self-reported health with their objective biomarker measures from 1998 to 2004 found that self-reported health was not as strongly associated with biomarker indicators of health among lower-income groups, indicating systematic differences in self-assessments of health status by income (Dowd & Zajacova, 2010). Other characteristics including age, gender, education, and other aspects of social status may also affect who an individual considers as belonging in their reference group when assessing their own overall health (Dowd & Zajacova, 2010; Lindeboom & van Doorslaer, 2004). Varying reference groups could influence the comparability of self-reports of overall health across societies or across groups within societies.

In response to these problems, some researchers have turned to other measures of well-being thought not to depend on judgments and reference groups, such as reports of specific symptoms, like flu symptoms, joint pain, or diarrhea. However, even the likelihood of acknowledging a specific symptom as abnormal or salient could vary for more and less-advantaged groups. Health disparities are likely to be underestimated if less-advantaged individuals report lower levels of symptoms or limitations than more-advantaged individuals when they have the same observable level of functioning (Bago D'Uva et al., 2008; Murray et al., 2002). Moreover, without appropriate attention to context and cultural characteristics of groups being compared, researchers may not measure aspects of health and functioning salient to more marginalized or unfamiliar groups, potentially mischaracterizing dimensions of inequality that study subjects themselves would prioritize (Hahn, 1999).

Finally, language or cultural factors influence respondents' interpretations of questions about their overall health or symptoms and their responses. When comparing the health of different groups using self-reported survey measures, researchers frequently assume that all respondents share definitions of that aspect of health or health overall, and that survey items are read and understood similarly by members of different linguistic and/or ethnic groups (Staniszewska et al., 1999). These assumptions may be unwarranted, such that estimates of health disparities may not measure differences in the underlying health indicator under study, but are driven by varying understandings or interpretations of the question or even differential comprehension.

Strategies for addressing the problem

Recognizing these problems with measures of self-rated health or symptoms, some researchers have pursued harmonization of instruments, or used vignettes and other surveys tools to try to improve data quality, while others have used statistical analytic solutions. While cultural factors clearly are essential to understanding the generation of health and health disparities, reducing misunderstanding and misinterpretation in the measurement of plausibly universal indicators of health status is a goal of many investigators interested in comparison across diverse groups. Data harmonization projects have translated the languages of survey instruments, but more effort may be needed to contextualize the intended comparisons and achieve cross-cultural equivalence of health indicators and the way they are collected (Hahn, 1999; Hunt & Bhopal, 2004).

In some studies, respondents are presented with several vignettes describing fictitious individuals and asked to rate those individuals' health, and researchers use these responses to identify reporting differences across groups. For example, Peracchi and Rossetti's (2012) study using data from a 2004 cross-national survey of 11 European countries showed that the use of vignettes can narrow systematic reporting differences on the basis of gender. Respondents were first asked to evaluate their own position on a scaled indicator in a given domain. They were all then asked to evaluate an identical vignette and indicate the health of the person described, using the same scale they used to rate their own position. Thus these "anchoring" vignettes are used to identify systematic individual or group-level variation in subjective thresholds used when people make self-assessments of overall health or other subjective indicators (Peracchi & Rossetti, 2012: 514).

Another approach for assessing the importance of reporting differences is comparing individuals' self-rated health with reports they make about their own specific health conditions to create an "adjustment factor" for the self-rated overall health measure. One study compared self-reported health and other reports about health conditions (such as BMI, treatment for depression, grip strength, and other quasi-objective health indicators) for respondents from ten European countries to address bias arising from different understandings of health states and different reference levels of health occurring across nations (Jurges, 2007). This study found that Danes and Swedes tend to over-rate their overall health based on their burden of specific conditions, while Germans under-rate their health, and confirms that some of the variation in self-reported health is due to differences in reporting styles across societies (Jurges, 2007). Using self-reports about diagnosed conditions or symptoms to adjust ratings of overall health could still be a limited solution, to the extent that reports of specific health statuses suffer from diagnosis bias or that reports of symptoms are affected by differential reference group issues.

Another approach designed to circumvent some of the problems of measures of self-rated overall health is to ask survey respondents about particular abilities or limitations while using very specific language to define these activities that most individuals need to perform. Activities of daily living (ADLs) items measure health by asking about level of ability in bathing, dressing, toileting, and being mobile indoors or on stairs (Katz et al., 1970). While these straightforward items are likely to reduce the influence of reporting differences across

groups, such scales are intended to measure levels of disability, a relatively nonspecific health status, and one best suited for older populations.

Alternatively, some researchers use statistical modeling to address individual or society-wide differences in reporting about overall health, symptoms or functioning. For example, “multi-method measurement models” (Eid et al., 2006), use multiple indicators (i.e., survey items about overall health, different symptoms, and health conditions) and generally apply structural equation modeling to identify a more comparable latent construct of health that reduces the influence of variation in reporting on different items. To directly model differences in styles of reporting associated with particular societies or groups, multilevel models with cross-level interactions between individual-level health reports and community or societal-level indicators may also be useful.

DISCUSSION

In this study we catalogued key challenges to health measurement that face health disparities research comparing across diverse societies and within unequal societies. While there are likely others, we focused on four that could affect a variety of comparisons: different distributions of disease, variation in the availability and quality of vital events and census data, variation in the comparability self-reports on specific health conditions due to diagnosis bias or diagnosis avoidance, and lack of comparability of self-reported overall health or symptoms measures. We discussed how these problems generally lead to underestimates of health disparities, but for some kinds of comparisons, they could also inflate estimates of disparities.

It is useful to illustrate how these different challenges might converge when considering a single focal health indicator in a particular context. A good example is infant mortality, an indicator commonly used in health disparities research because it accurately taps material deprivation and is widely available. The quality of South African vital events data on infant births and deaths has long varied widely by population groups defined by race, and these differences map onto vast and longstanding differences in socioeconomic standing. Records for black South Africans (often identified in South Africa as “Africans”) have suffered from errors of coverage and content to a much greater extent than those for the most-advantaged white group. Deaths of non-white South African infants have historically been underreported, especially deaths of African infants, leading to artificially-low estimates of disparities in the overall infant mortality rate. However, lower birth registration coverage for non-whites (which drives the denominator of the rate) or differential reporting on surveys of births or deaths can also result in overestimates of disparities between less- and more-advantaged South Africans (Nannan et al., 2007). Additionally, cause-of-death data has been less accurate for poorer areas and their non-white residents (Bradshaw & Steyn, 2002), and deaths due to HIV/AIDS, a highly stigmatized condition, have been greatly underreported or not identified (Birnbau et al., 2011). These issues of accuracy in death reporting affect estimates of disparities in the cause-of-death profiles of white and non-white infants. Most centrally, rural African infants perish at much higher rates from preventable, post-neonatal causes than more socioeconomically-advantaged white infants and those born to urban residents (Andersson & Marks, 1988). These issues illustrate three of the challenges we

discussed in our review. They are likely to complicate the use of infant mortality data within the deeply racially-stratified context of South Africa, but also across societies at different levels of economic development and with differing vital events recording infrastructures and disease profiles.

Several issues were beyond the scope of this review, but are related to the ones discussed here and deserve further attention. First, we have focused on problems in measuring disparities when using cross-sectional data. These problems could be compounded for researchers interested in studying how health disparities change over time. Quality and coverage of health measurement could be changing at different rates across groups and societies, potentially clouding patterns of convergence or divergence in health disparities. Second, measurement of the predictors of health, including socioeconomic position or discrimination, may vary in accuracy and reliability across groups and contexts (Stewart & Nápoles-Springer, 2003), further complicating the goal to study evolving disparities in health along these axes.

Our review suggests several clear recommendations. The brief examples we outline in this paper are useful, but it is important to remember that the choice of comparison societies or social groups within societies will determine which of these challenges are most salient for a given study and how they play out. Also, challenges will likely be greatest for comparisons where the social and economic distance between societies or groups is large. Since there is no single adequate indicator or strategy, researchers need to design and report on health disparities studies that make clear their choices and assumptions. Moreover, choices about what measures to use should also be viewed not only as a problem of measurement but as a problem of operationalization. Disparities researchers are trying to measure a vastly multidimensional concept—health—with a wide possible range of indicators, and their varying levels of presence, quality, comparability, validity and reliability must be considered. Researchers should be aware of possible sources of bias and discuss the directions in which their estimates of disparities are likely to be biased, given the specific comparison groups, social contexts, health outcome or outcomes, and sources of data. Application of adjustments for differences in data availability, quality, or reporting styles could provide bounds on estimates in some cases, whether by using vignettes, simulations, or estimates drawn from other studies of similar contexts or health conditions. Collecting biomarker data is costly, but given the seriousness of diagnosis and reporting biases, efforts to include such measures are laudable when such indicators are appropriate.

An overarching theme embedded in this commentary is the importance of contextualizing not only the measures of health available for any given comparison – from vital statistics to self-reports of chronic conditions – but also the disparities in health that are the focus of measurement. As Krieger (2011) argues, both health and its measurement are deeply embedded in specific socioeconomic, political and historical contexts. Ignoring ways that the generation and measurement of health in specific societies are influenced by resilient systems of power and privilege may complicate attempts to improve measurement or health itself, or even render them ineffective. Though it is beyond the scope of this commentary to conduct the necessary political, economic, historical and cultural analysis of the biographies

of the measures available (or unavailable) for any particular comparison, health disparities research would be strengthened by such a sensitivity to context.

Our call for sensitivity to these measurement challenges must also acknowledge that the use of health measures is always politicized, whether they are used to advocate for underserved groups facing inequitable health burdens, or to legitimate an existing system of power and privilege (Etches et al., 2006). In this commentary we propose that effectively addressing the poorer health of disadvantaged groups demands critical appraisal and careful and informed use of existing measures of health. Clearly, improvements in available data are also warranted. Continuing to strengthen the quality of existing data collection systems and maintaining consistent quality standards for socially-advantaged and disadvantaged groups is a goal of many administrators of vital statistics records, censuses, surveys, and other major data collections, and one worth supporting despite the associated challenges. Further exploration of the viability of novel data sources for health disparities research is also needed. All of these efforts could improve measurement and interpretation of health disparities in a highly unequal and rapidly changing world, and if used judiciously, may aid efforts to close gaps or increase population health overall.

Acknowledgments

This study was supported by core funding from Eunice Kennedy Shriver National Institute of Child Health and Human Development grant R24 HD041028 to the Population Studies Center, University of Michigan. We would like to thank Jessica Hoel and Apoorva Aekka for excellent research assistance and Donald J. Treiman and Barbara Anderson for fruitful discussions about the ideas in this paper.

REFERENCES

- Anand S, Hanson K. Disability-adjusted life years: a critical review. *Journal of Health Economics*. 1997; 16:685–702. [PubMed: 10176779]
- Andersson N, Marks S. Apartheid and health in the 1980s. *Social Science & Medicine*. 1988; 27:667–681. [PubMed: 3067361]
- Ayles H, Schaap A, Nota A, Sismanidis C, De Haas P, Muyoyeta M, et al. Prevalence of Tuberculosis, HIV and Respiratory Symptoms in Two Zambian Communities: Implications for Tuberculosis Control in the Era of HIV. *PLoS ONE*. 2009; 4
- Bago D’Uva T, O’Donnell O, van Doorslaer E. Differential health reporting by educational level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology*. 2008; 37:1375–1383. [PubMed: 18676985]
- Bardage C, Pluijm SMF, Pedersen NL, Deeg DJH, Jylhä M, Noale M, et al. Self-rated health among older adults: a cross-national comparison. *European Journal of Ageing*. 2005; 2:149–158.
- Birnbaum JK, Murray CJ, Lozano R. Exposing misclassified HIV/AIDS deaths in South Africa. *Bulletin of the World Health Organization*. 2011; 89:278–285. [PubMed: 21479092]
- Blumberg S, Foster E, Frasier A. Design and operation of the National Survey of Children’s Health, 2007. National Center for Health and Statistics. 2012
- Boerma JT, Stansfield SK. Health Statistics Now: Are We Making the Right Investments? *The Lancet*. 2007; 369:779–786.
- Bradshaw, D.; Steyn, K., editors. *Poverty and Chronic Diseases in South Africa: Technical Report* 2001. South Africa Medical Research Council; Cape Town, South Africa: 2002.
- Bramley D, Hebert P, Tuzzio L, Chassin M. Disparities in Indigenous Health: A Cross-Country Comparison Between New Zealand and the United States. *American Journal of Public Health*. 2005; 95:844–850. [PubMed: 15855464]

- Braveman P. HEALTH DISPARITIES AND HEALTH EQUITY: Concepts and Measurement. *Annual Review of Public Health*. 2006; 27:167–194.
- Braveman P, Cubbin C, Egerter S, Williams DR, Pamuk E. Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us. *American Journal of Public Health*. 2010; 100:S186–S196. [PubMed: 20147693]
- Byass P, Kahn K, Fottrell E, Mee P, Collinson MA, Tollman SM. Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. *Population Health Metrics*. 2011; 9:46. [PubMed: 21819601]
- Conteh L, Engels T, Molyneux DH. Socioeconomic aspects of neglected tropical diseases. *The Lancet*. 2010; 375:239–247.
- Crimmins, E.; Guyer, H.; Langa, K.; Ofstedal, MB.; Wallace, R.; Weir, D. HRS Documentation Report: Documentation of Biomarkers in the Health and Retirement Study. Survey Research Center, University of Michigan; Ann Arbor MI: 2008.
- Dahlgren G, Whitehead M. A Framework for Assessing Health Systems From the Public's Perspective: The ALPS Approach. *International Journal of Health Services*. 2007; 37:363–378. [PubMed: 17665729]
- DeSalvo K, Fan V, McDonnell M, Fihn S. Predicting Mortality and Healthcare Utilization with a Single Question. *Health Services Research*. 2005; 40:1234–1246. [PubMed: 16033502]
- DFID. Better health for poor people. Department For International Development; London, UK: 2000.
- Dowd JB, Zajacova A. Does Self-Rated Health Mean the Same Thing Across Socioeconomic Groups? Evidence from Biomarker Data. *Annals of Epidemiology*. 2010; 20:743–749.
- Dwyer L. Collecting medication data in the 2004 National Nursing Home Survey. *Vital Health Stat: National Center for Health Statistics*. 2009
- Eid, M.; Lischetzke, T.; Nussbeck, F. Structural Equation models for multitrait multimethod data. In: Eid, M.; Diener, E., editors. *Handbook of multimethod measurement in psychology*. American Psychological Association; Washington DC: 2006.
- Epstein A. Health Care in America--Still Too Separate, Not Yet Equal. *New England Journal of Medicine*. 2004; 351:603–605. [PubMed: 15295055]
- Etches V, Frank J, Ruggiero ED, Manuel D. MEASURING POPULATION HEALTH: A Review of Indicators. *Annual Review of Public Health*. 2006; 27:29–55.
- Felix-Aaron K, Moy E, Kang M, Patel M, Chesley F, Clancy C. Variation in the Quality of Men's Health care by Race/Ethnicity and Social Class. *Medical Care*. 2006; 43:172–81. [PubMed: 15746594]
- Frenk J, Bobadilla JL, Sepulveda J, Cervantes ML. Health transition in middle-income countries: new challenges for health care. *Health Policy and Planning*. 1989; 4:29–39.
- Gaur S, Whitley-Williams P, Flash C, Jimenez E, Petrova A. Disparity in Hospital Utilization of Rapid-HIV-1 Testing for Women in Labor with Undocumented HIV Status. *Maternal and Child Health Journal*. 2010; 14:268–273. [PubMed: 19259801]
- Hahn, RA. Anthropology and the Enhancement of Public Health Practice. In: Hahn, RA., editor. *Anthropology in Public Health: Bridging Differences in Culture and Society*. Oxford University Press; New York: 1999. p. 3-26.
- Hajjar I, Kotchen T. Trends in Prevalence, Awareness, Treatment, and Control of Hypertension in the United States, 1988-2000. *Journal of American Medical Association*. 2003; 290:199–206.
- Hosler AS, Berberian EL, Spence MM, Hoffman DP. Outcome and Cost of a Statewide Diabetes Screening and Awareness Initiative in New York. *Journal of Public Health Management and Practice*. 2005; 11:59–64. [PubMed: 15692294]
- Hunt SM, Bhopal R. Self report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture. *Journal of Epidemiology and Community Health*. 2004; 58:618–622. [PubMed: 15194728]
- Idler E, Benyamini Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior*. 1997; 38:21–37. [PubMed: 9097506]
- Johnston D, Propper C, Shields M. Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*. 2009; 28:540–552. [PubMed: 19406496]

- Jurges H. True Health vs. Response Styles: Exploring Cross-Country Differences in Self-Reporting Health. *Health Economics*. 2007; 16:163–178.
- Kahn K, Garenne ML, Collinson MA, Tollman SM. Mortality trends in a new South Africa: Hard to make a fresh start. *Scandinavian Journal of Public Health*. 2007; 35:26–34. [PubMed: 17676500]
- Katz S, Downs T, Cash H, Grotz R. Progress in Development of the Index of ADL. *The Gerontologist*. 1970; 10:20–30. [PubMed: 5420677]
- Kellerman S, Lehman S, Lansky A, Stevens M, Hecht F, Bindman A, et al. HIV Testing Within At-Risk Populations in the United States and the Reasons for Seeking or Avoiding HIV Testing. *Journal of Acquired Immune Deficiency Syndrome*. 2002; 31:202–210.
- Kranzer K, van Schaik N, Karmue U, Middlekoop K, Sebastian E, Lawn S, et al. High Prevalence of Self-Reported Undiagnosed HIV despite High Coverage of HIV Testing: A Cross-Sectional Population Based Sero-Survey in South Africa. *PLoS ONE*. 2011; 6
- Krieger, N. *Epidemiology and the People's Health: Theory and Context*. Oxford University Press; New York: 2011.
- Lima-Costa MF, De Oliveira C, Macinko J, Marmot M. Socioeconomic Inequalities in Health in Older Adults in Brazil and England. *American Journal of Public Health*. 2012; 102:1535–1541. [PubMed: 22698020]
- Lindeboom M, van Doorslaer E. Cut-point shift and index shift in self-reported health. *Journal of Health Economics*. 2004; 23:1083–1099. [PubMed: 15556237]
- Mackenbach JP, Looman CWN, van der Meer JBW. Differences in the Misreporting of Chronic Conditions, by Level of Education: The Effect on Inequalities in Prevalence Rates. *American Journal of Public Health*. 1996; 86:706–711. MD PHD. MSc. MD. [PubMed: 8629723]
- Marmot M, Friel S, Bell R, Houweling TA, Taylor S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet*. 2008; 372:1661–1669. [PubMed: 18994664]
- Mont D. Measuring health and disability. *Lancet*. 2007; 369:1658–1683. [PubMed: 17499607]
- Murray C, Frenk J. Health metrics and evaluation: strengthening the science. *The Lancet*. 2008; 371:1191–1199.
- Murray C, Tandon A, Salomon JA, Mathers CD, Sadan R. Cross-population comparability of evidence for health policy. *Global Programme on Evidence for Health Policy: World Health Organization*. 2002
- Nannan N, Timaeus IM, Laubscher R, Bradshaw D. Levels and differentials in childhood mortality in South Africa, 1977–1998. *Journal of Biosocial Science*. 2007; 39:613–632. [PubMed: 17107633]
- Ojanuga D, Gilbert C. Women's access to health care in developing countries. *Social Science & Medicine*. 1992; 35:613–617. [PubMed: 1519115]
- Omran AR. The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *Milbank Memorial Fund Quarterly*. 1971; 49:509–538. [PubMed: 5155251]
- Peracchi F, Rossetti C. Heterogeneity in health responses and anchoring vignettes. *Empirical Economics*. 2012; 42:513–538.
- Preston SH, Elo IT, Rosenwaike I, Hill ME. African-American Mortality at Older Ages: Results of a Matching Study. *Demography*. 1996; 33:193–209. [PubMed: 8827165]
- Sankoh O, Byass P. The INDEPTH Network: filling vital gaps in global epidemiology. *International Journal of Epidemiology*. 2012; 41:579–588. [PubMed: 22798690]
- Schneider M, Bradshaw D, Steyn K, Norman R, Laubscher R. Poverty and non communicable diseases in South Africa. *Scandinavian Journal of Public Health*. 2009; 37:176–186. [PubMed: 19164428]
- Sen A. Health: Perception Versus Observation: Self Reported Morbidity Has Severe Limitations and Can Be Extremely Misleading. *British Medical Journal*. 2002; 324:860–861. [PubMed: 11950717]
- Siegel, JS.; Swanson, DA., editors. *The Methods and Materials of Demography*. Second Edition. Elsevier Academic Press; San Diego and London: 2004.
- Staniszewska S, Ahmed L, Jenkinson C. The conceptual validity and appropriateness of using health-related quality of life measures with minority ethnic groups. *Ethnicity and Health*. 1999; 4:51–63. [PubMed: 10887462]

- Stewart AL, Nápoles-Springer AM. Advancing Health Disparities Research: Can We Afford to Ignore Measurement Issues? *Medical Care*. 2003; 41:1207–1220. [PubMed: 14583684]
- Steyn, K.; Schneider, M. Poverty and Chronic Diseases in South Africa. In: Bradshaw, D.; Steyn, K., editors. *Poverty and Chronic Diseases in South Africa: Technical Report*. South Africa Medical Research Council; Cape Town, South Africa: 2001. p. 1-14.
- Tabana H, Doherty T, Swanevelder S, Lombard C, Jackson D, Zembe W, et al. Knowledge of HIV status prior to a community HIV counseling and testing intervention in a rural district of south Africa: results of a community based survey. *BMC Infection Diseases*. 2012; 12:1–8.
- Van Dyk A, Van Dyk P. “What is the point of knowing?” : psychosocial barriers to HIV / AIDS voluntary counselling and testing programmes in South Africa. *South African Journal of Psychology*. 2003; 33:118–125.
- Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients’ knowledge of their chronic disease. A study of patients with hypertension and diabetes. *Arch Intern Med*. 1998; 158:166–172.
- Willson, Andrea E.; Shuey, Kim M.; Elder, Glen H.J. Cumulative Advantage Processes as Mechanisms of Inequality in Life Course Health. *American Journal of Sociology*. 2007; 112:1886–1924.
- World Health Organization. *Global Health Risks: Mortality and burden of disease attributable to selected major risks*. Switzerland; Geneva: 2009.
- Zhang AY, Snowden LR, Sue S. Differences between Asian and White Americans’ help seeking and utilization patterns in the Los Angeles area. *Journal of Community Psychology*. 1998; 26:317–326.

RESEARCH HIGHLIGHTS

- We describe some key challenges to health measurement for health disparities research
- When disease profiles differ across groups, some measures misrepresent disparity
- Multiple types of bias could arise when comparing any measure of health across groups