

PERSPECTIVES

The Rise, Critique and Persistence of the DALY in Global Health

Rachel Parks

Princeton University, Princeton, NJ, USA

A barrage of criticism from global health scholars in disciplines ranging from economics to anthropology immediately followed the introduction of the disability-adjusted life year, or DALY, in 1994. Despite demonstrated flaws in its justification and design, the DALY is still in wide use in the field of global health, promoted by scholars as well as by major publications like *The Lancet* and by funding agencies like the Gates Foundation. Two case studies first pointed out flaws in the mid-1990s that are still being overlooked in current research projects. The DALY has persisted while the power structure of global health has changed from the political to the economic and biomedical, and power (and money) have become concentrated in the hands of a few individuals. This change in regime is the reason that the DALY has persisted despite its flaws, but it also has positive implications for the future. Because of the extreme concentration of power in global health today, it would not be impossible to uproot the DALY, even though it is so widely used. If research dollars were devoted to the development of a better metric, only a handful of leaders would need to be convinced of its value in order for it to take over. The elimination of the DALY and development of a replacement is therefore not only necessary, as has been pointed out for almost twenty years, but also feasible, due to more recent changes in the political structure of global health. A new, two-part metric is proposed that would address the most common critiques of the DALY while still providing numerical guidance for health policy decision-making.

INTRODUCTION

What is the DALY?

The disability-adjusted life year, or DALY, is a metric designed to quantitatively measure the impact of various diseases and conditions on the productivity and well-being of people through a combination of mortality and morbidity estimates. As a numerical value that can be compared across nations, its use has become widespread in policymaking, academia and nonprofit work. Although the DALY has continually grown in popularity, it has also been contested since its inception, notably in Sudhir Anand and Kara Hanson's 1997 paper, and by anthropologists such as Vincanne Adams.¹ Even leaders from "hard" sciences, such as Director of the National Institutes of Health (NIH) Francis Collins, have called DALYs and similar metrics like the quality-adjusted life year "only partially successful in providing the kind of information that policy-makers need," and urged the NIH to fund the "development and application of more rigorous models."² The DALY has been evaluated by members of various disciplines since the mid-1990s, and this paper will outline their arguments and provide examples of how the drawbacks of DALYs can at times impede the goals of global health. Despite its failings, the DALY is still widely used by researchers, and the following analysis will attempt to illuminate how the current power structure in global health is preventing viable alternatives from being developed.

Christopher Murray publicly introduced the DALY in 1994, in an article published by the World Health Organization (WHO) entitled "Quantifying the burden of disease: the technical basis for disability-adjusted life years."³ In the paper, he explained his justifications for creating the DALY and described the technical details of the metric. His goal was to open the "black box" of policymakers' values in public health by attaching numerical values to various health conditions and disabilities in order to create a metric that would combine both mortality and morbidity into a single value.³ This "black box" referred to the "wide variation in the implied value of saving a life" evident in different pieces of public safety legislation.³ The DALY addressed this variation by creating a standard way of calculating the value of peoples' lives and, therefore, the amount of money that should be spent to help them.

Calculating DALYs involves several steps. First, potential years

of healthy life lost are calculated using one of several life-expectancy measures, and, in the case of "non-fatal health outcomes," are then multiplied by a "disability weight." These weights are values between zero and one, with zero representing full health and one representing death, determined by "an independent group of experts."³ The years of life are also adjusted by the presumed differential societal value of individuals at different ages, with the negative impact of a death peaking around age 12 and reaching negligible impact at age 100 (Figure 1).³ Therefore, individual deaths contribute some number of DALYs that varies depending on age and disability status. In practice, this means that the deaths of individuals who are old, sick or disabled contribute less to the estimated burden of disease. Many researchers have taken this unequal valuing of life to be an ethical problem that could someday manifest itself as a practical one, as will be discussed below. Apart from these ethical issues, many scholars see problems with the actual statistical uses of the DALY metric, as documented below.

What should DALYs measure?

Jeffery Hammer, an economist, argues that the goal that DALYs help to achieve is not the right one for policymakers to be pursuing. For Hammer, a debate about the specifics of the DALY is less fundamental than the debate about which economic model should be used for building health systems. The three main categories of goals for health systems, he writes, are "improving aggregate health status," "improving equity and reducing poverty" and "improving individual welfare."⁴ Of the three goals that Hammer describes, he emphasizes that DALYs or cost-effectiveness analysis can only reasonably address the goal of improving aggregate health status. When people argue against the use of the DALY, he says, they are really disputing that specific economic model, not the calculation itself. "Discussions of means are often confused by what are really disagreements about ends," he writes.⁴

Hammer also questions the means by which policymakers attempt to address the goal of improving aggregate health status, which they support by using the DALY. Improving aggregate health involves "allocating limited resources to the provision of treatments for those diseases which have the highest health impact per dollar spent."⁴ But

Jack's Art's Gallery

*Custom framing · Custom Mirrors ·
10% off with CUID*

*2855 Broadway
(Between 110th and 111th Streets)
New York, NY 10025*

Hours:

*Monday - Saturday: 10am to 8pm
Sunday: 11am to 6pm
Friday closed: 1:30 to 2:30*

Phone: (212) 749 - 5554

www.jacksartframing.com

Email: contact@jacksartframing.com



Get it @ "theMarket"

Locations:

2840 Broadway
(110th St. & Broadway)

2589 Broadway
(97th St & 98th St)

2171 Broadway
(76th St & 77th St)

77 7th Avenue
(14th St & 15th St)

The highest quality foods and hard-to-find products supported by friendly, personalized, efficient service

"Morton Williams, founded in 1946, is a family-owned and operated food retailer with eleven stores in the New York City Metropolitan area. Each store is designed to reflect the needs of the individual neighborhood."

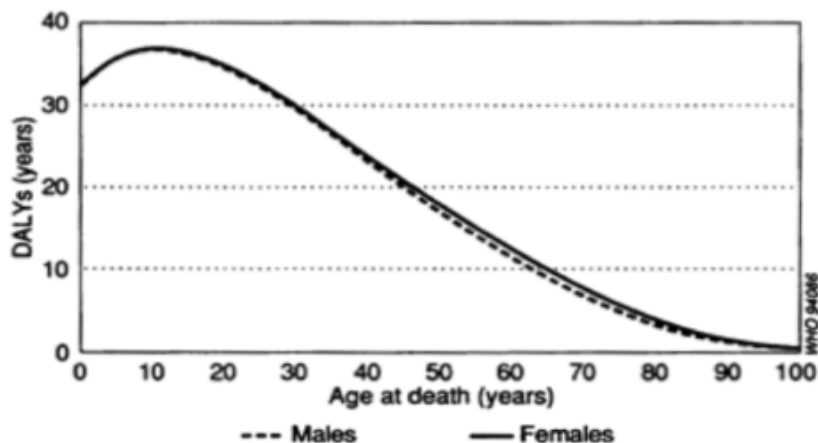


Address:

*2941 Broadway (115th Street)
New York, NY 10025*

Phone: 212-666-4190

Open 24 hours a day

Figure 1: Curve showing age-weighting by gender in DALY calculations³

if the state or other funding agencies are entering the business of providing health care, Hammer says, they should be focusing on “projects which yield the greatest improvement in the measure of health status chosen relative to what would happen if the Ministry did not do them” (author’s emphasis).⁴ That is, the government could provide a service that would yield a benefit, but if the service would have otherwise been provided by the private sector, then not all of the benefit could be attributed to the government’s action. Therefore, it is not always best to devote the most resources to the health problem that yields the most DALYs. To Hammer, DALYs “make no sense” because they are implemented according to an economic model that ignores the way the world works in practice.⁵ He sees this lack of awareness as an ethical problem because the uninformed decisions it leads to can potentially affect millions, if not billions, of people.

Michael Reich, another economist, sees as the main flaw of the DALY its goal of being a “double metric,” meaning it seeks to increase efficiency as well as equity.⁶ Reich disapproves of the DALY’s use as a means to achieve two of Hammer’s three goals (improving aggregate health status and improving equity), rather than only the one that Hammer argues that it is appropriate for (improving health status). The problem is that these two goals are not always aligned; Reich criticizes the 1993 World Development Report (WDR), for which the DALY was designed, for not specifying what to do “when cost-effectiveness and equity are in conflict.”⁶ Reich also sees ethical issues embedded in the use of the DALY, yet in the end, he does not fully reject it, because it “stands head and shoulders above all others [metrics used in the WDR].”⁶ He believes that combinatory measures of health are helpful and necessary, and that the problems with the DALY are less severe than problems with other proposed metrics. Reich might explain the continued use of the DALY in terms of convenience; it was invented for the World Health Organization, made readily available, and nothing better has emerged since.

Arguments against the DALY

Anand and Hanson, two economists who wrote an early, strongly-stated critique of Murray’s work, frame the problem in a way similar to Reich’s paper. Their primary critique of the DALY is that it attempts both to measure the global burden of disease and to guide the allocation of resources.¹ In their paper, they explain how Murray failed to meet either of these goals individually and therefore why the DALY as a whole does not work.

First, according to Anand and Hanson, the DALY failed to provide a measure for resource allocation because it did not ac-

count for differentials in resource availability. This idea parallels Hammer’s critique that governments ignore the contributions of the private sector (or lack thereof) when determining their own impact on health. Murray justifies his weighting technique with the example that “the premature death of a 40-year-old woman should contribute equally to estimates of the global burden of disease irrespective of whether she lives in the slums of Bogota or a wealthy suburb of Boston.”³ But Becker and her colleagues challenge readers to wonder “whether these two deaths are really alike,” since the individuals have different resources available and might be fulfilling different roles in their communities and families.⁷ Anand and Hanson also argue that since DALYs are based on baseline measurements from wealthy countries, the differential found between these populations and those of developing countries measures the “burden of disease and underdevelopment, and not that of disease alone” (authors’ emphasis).¹ Both on an individual and a national scale, the DALY fails to account for differences in resources.

The DALY also fails at its other goal: that of measurement of disease burden, for both statistical and ethical reasons. In their arguments against the technical details of the DALY, Anand and Hanson were among the many who found fault with the fact that the weighting system was established by a “group of independent experts,” stating that there was no way to assess “the statistical or scientific basis for selecting the weights and, thus...their validity.”¹ Murray set out to eliminate the “black box” of policy decision-making by standardizing it. However, he chose to standardize

the DALY by relegating the task of weighting to an unnamed group of “experts” and failed to account for the problem he had originally pointed out.

Because the DALY is a statistical measure, technical arguments against it are important, but what people have written about the ethical problems is also compelling. Anand and Hanson

Disabled activists make a philosophical argument that their lives should be valued equally to those of people with no disabilities.

argue philosophically against the very idea of devaluing lives, pointing out that “by definition, DALYs are a ‘bad’ which should be minimized,” even though “more of a ‘life-year (even ‘adjusted’) should be a ‘good’, which should be maximized and not minimized.”¹ Discounting life in general is problematic, but especially so for people with life-long disabilities; disabled activists make a philosophical argument that their lives should be valued equally to those of people with no disabilities (see, for example, the work of Dr. Adrienne Asch). The World Health Organization was criticized for seemingly devaluing the lives of disabled people, and the WHO responded to this criticism by making the language explaining devaluation in the new version of its protocol even more explicit.⁸ Becker et al. ask whether the stigma of disability should be factored into the DALY algorithm in order to address this problem.⁷ However, devaluing the life of someone with a congenital disability is itself a form of discrimination. Factoring stigma into the algorithm would result in the devaluation of the lives of those with disabilities even further, as well as validating the stigma against them. Especially with congenital disabilities, which are usually not curable, decision-making based on cost-effectiveness analysis would result in widespread defunding of health services for this population. Arnesen and Nord proposed that this problem is due to the “functional capacity,” or economic value, of humans being confused with the actual value of their lives.⁸ In order for a metric to accomplish what the DALY attempts, it is necessary to combine all these factors. Measuring the burden of disease and resource allocation together requires conflating economics with

health, and therefore even a better-calculated metric would still face the ethical conundrum of putting a price on life.

Reidpath et al. similarly critiqued the ethical implications of the disability weights in the DALY, but they were especially critical of the way later iterations of the DALY, designed to address these issues, failed to make the measure more equitable. Major critics in the 1990s objected to the original DALY's emphasis on the ability to perform activities associated with "normal" human life, and ignored "the social, cultural or environmental context of the condition."⁹ In response, Murray and his team asked judges of disability weights to consider the "average handicap" associated with the social situation of people with certain disabilities, including stigma and other cultural factors. But Reidpath and his colleagues argued that, because of regional differences, this "average" handicap is just as useless as not taking culture into account at all. The authors found it inappropriate that, for example, the "same disability weight would be used in the calculation of the DALYs associated with epilepsy in Bogota, Beijing, New York and Newcastle."⁹ The authors insisted, however, that even with regional differences taken into account, the measure still would not be equitable, returning to the argument about differential resources in disease treatment that Anand and Hanson referred to as the "burden of underdevelopment."⁹

CASE STUDIES

A wide array of convincing economic and ethical arguments against the DALY were deployed well before the year 2000. One might expect, given this history, that the DALY would have been abandoned, either along with all other summary metrics or in favor of a less problematic one. Yet in June 2013, *The Lancet*, one of the world's leading medical journals, published a special issue devoted to the topic of health metrics entitled "Global Health Metrics & Evaluation: Data, Debates, Directions."¹⁰ Out of the first twenty conference abstracts presented, five of them used DALYs, and most of the others measured variables that the DALY could not be applied to, such as malarial parasite density, which they did not relate to the burden of disease.¹⁰ Case studies, presented below, show that researchers such as those published in *The Lancet* often fail to take into account critiques of the metric upon which they chose to base their policy recommendations.

Case Study 1: *The Lancet* and Disease Burden in Kenya

Anand and Hanson pointed out the problematic nature of having an unspecified group of "experts" decide on disease weights. The following example provides more specificity, but brings up even more questions about the decision-making process. One of the abstracts in *The Lancet's* special issue on metrics summarized a presentation about the authors' attempt to establish disability weights (DWs) for pediatric congenital anomalies by surveying health professionals in Canada and Kenya.¹¹ Out of 15 "health states" measured, two were significantly different between the two countries with a p-value below 0.0001. For one of them, cleft lip and palate, Canadians ranked the disability at 0.25, about the same as inflammatory heart disease or a pelvic fracture, while Kenyans ranked it as important as a lower arm fracture or malaria.^{11,12} Despite this, the authors declared in their conclusion that they were successfully able to establish new disability weights because "DWs do not appear to differ significantly across cultural contexts."¹¹ Faced with such a large, statistically significant disparity for two of the syndromes, the authors simply averaged the Canadian and Kenyan values for their final report, not being able

to come up with any better method.¹¹ This is problematic and its risks can be demonstrated anecdotally in the case of Dr. Poenaru, where upon moving to Kenya to practice surgery, "cleft lip and palate repair—a plastic surgeon's domain—has become his bread and butter."¹³ The fact that he considers the burden of cleft palate greater than his patients do could cause him to perform more surgeries than they think are necessary, thereby directing limited resources towards that problem and subjecting patients to the unavoidable risks of surgery for reasons that they might not consider adequate. The "black box" of decision-making that Murray complained about is certainly opened by the DALY, but the "value choices" he talks about are ensconced by the introduction of his metric.³

Case Study 2: Disease Spending in Tanzania

Becker et al. provide an analytical critique of the DALY, as described above, but they also cite a study of health spending in Tanzania as an example of effective use of DALYs to align health care spending with needs. Before the study, the share of expenditures on each of various programs was either above or below the share of DALYs it was calculated to contribute, and afterwards, the levels were almost perfectly aligned. The budget for two areas (the Expanded Program on Immunization and TB DOTS, or short-course directly-observed tuberculosis treatment) fell significantly.⁷ The reason that these programs were calculated to comprise such a small proportion of the burden of disease illustrates another key flaw in the DALY; it discounts future lives at a 3% rate per year compared to current lives.¹

Anand and Hanson extrapolate this discount rate to its absurd conclusion, namely that there is "a 50% chance that the world will end in 23.4 years," which is when future lives are discounted all the way to zero.¹ So the "success" of this program, therefore, was mostly predicated on the fact that it severely discounted the lives of future generations, and the funding changes it guided make future patients more likely to face both re-emergent infectious diseases that had previously been controlled by immunizations, as well as multidrug-resistant tu-

berculosis. Even critical scholars in global health can sometimes overlook problems related to the metrics they use. Becker et al. list Anand and Hanson's paper as a "suggested reading," but fail to account for one of its most convincing arguments in their analysis.⁷ Admittedly, the authors do lament the "anemic response to multidrug-resistant tuberculosis," but they attribute it to the problems with using cost-effectiveness analysis in general, rather than seeing it as an easily avoidable pitfall associated specifically with the DALY.⁷

DALYs and Lack of Data

Even if one assumes that the DALY, having now been in use for two decades, is a permanent fixture in the study of global health, various people have proposed ways of improving it. Jeff Hammer's main critique of the DALY does not focus on its embedded value judgments, but in the lack of data behind its general use.¹⁴ Becker and colleagues describe the extrapolation used in the World Development Report as requiring a "leap of faith."⁷ Cooper et al. pointed out that the reported numbers for most of the 48 countries in sub-Saharan Africa were based on records from South Africa alone, which account for only 1% of the sub-Saharan population.¹⁵ This extrapolation is simply based on GDP, and is unbelievably common; Hammer pointed out that each actual observation in the Global Burden of Disease report was

The most fundamental issue with the DALY, namely that it tries both to measure the burden of disease and to direct funding, could be addressed by introducing not one, but two new metrics; one for burden, and one for need.

used in models to create 1,500 additional entries.¹⁴ Those who work on estimating disease burden should prioritize the collection (rather than extrapolation) of “vital statistics”: records of birth, death, and changes in marital status, usually reported by country. However, both the collection and the coding of cause-of-death are often questionable. Arthur Kleinman, a psychiatrist and medical anthropologist, states this very strongly: “mortality rates are social fabrications that are based upon often seriously inadequate sets of data of questionable accuracy.”¹⁶ Jeff Hammer cited a study by Veena Das in which she found that death by “broken heart,” interpreted as cardiovascular disease, was actually a death attributed to grief over the loss of a spouse.¹⁴

Areas of neglect in global health can be explained by lack of funding. The fact that better data are not being collected suggests that big donors are not interested in measuring health, but this is actually not the case. International funding organizations like WHO and AusAID are spending money to develop sophisticated toolkits to use to monitor country-wide collection of vital statistics; that is, to monitor the actual collection of data.¹⁷ The Gates Foundation is also a major funder of work in metrics as they strive to meet Bill Gates’s Grand Challenge #13: Develop Technologies that Permit Quantitative Assessment of Population Health Status.¹⁸ One of the recipients of this funding is the Institute for Health Metrics and Evaluation (IHME).

Explaining the Persistence of the DALY

Since the IHME is currently directed by Christopher Murray, the developer of the DALY, the institute likely holds a preference for studying combinatory metrics rather than actual morbidity and mortality rates. The organization, according to his Director’s Statement, has contributed to global health by inventing new tools for identifying causes of death, documenting global health expenditures, and “creating new ways of measuring health challenges,” including combinatory metrics like the DALY.¹⁹ The IHME’s website fails to reference performing any data collection. The organization’s principles state that they aim to base “measurements on...available data and objectively portray the uncertainty in measurements” and to “consult with the global health community” even though “consultation does not necessarily lead to consensus.”¹⁹ In essence, this means that the availability of data is more important than either its completeness or its importance to the actual communities being studied. The actual practice of the IHME could be different, but its rhetoric privileges mathematical calculations over human needs.

The persistence of the DALY comes from the way in which Murray, in his career, has navigated the change in power structure in the world of global health. He created the DALY for the World Health

Organization, but he moved on in 2003 and in 2007 became the director of the IHME, which is located in Seattle, Washington.¹⁹ Jeff Hammer described the way the center of power of global health has moved from Washington, D.C. to Seattle. The U.S. government, as the most powerful voting member of the World Health Organization, used to be the ruling power, but the Gates Foundation has since eclipsed it.¹⁶ It is not surprising that Gates, with his background in business, finds numerical computational models of disease appealing, and in fact the Gates Foundation underwrote many of the papers in *The Lancet* special issue on metrics.¹⁰ The DALY, then, has survived partly because it has received the stamp of approval of a few powerful actors.

Vincanne Adams, an anthropologist, discusses the role of metrics in the new political regimes of global health. Metrics, she says, can be seen as a good thing, since in a way they counteract political power, which is Murray’s original justification for the DALY.²⁰ However, she points to two problems: one, of finding a metric that can serve as a universal standard, and two, of the new kinds of sovereignty that new metrics will make possible.²⁰ The DALY clearly failed at serving as an appropriate universal standard, but it persisted because of the new, economically justified biomedical sovereignty that it helped to usher in, as the major source of power changed from politics to economics.

CONCLUSION

An Alternative to the DALY?

Francis Collins, who, like Gates, is in charge of one of the world’s most important funding sources of health research, is critical of the DALY. Although Collins speaks out against current metrics, he still encourages the use of summary statistics in general. Anthropologists may always critique efforts to summarize the human experience using numbers, but there will, for the foreseeable future, always be people in power in the field of global health who will want summary statistics of the burden of disease. Nonetheless, there are certainly criticisms of the manner in which the DALY does its job that are convincing even to people who find numbers more compelling stories, and which could be solved by introducing a new metric for disease calculation.

The most fundamental issue with the DALY, namely that it tries both to measure the burden of disease and to direct funding, could be addressed by introducing not one, but two new metrics; one for burden, and one for need. Reidpath et al. called for a “measure that included context,” which would “more closely reflect the realities of the burden of disease.”⁹ This metric for burden could be based, not on extrapolations from GDP, but on real measurements of morbidity and mortality, and the disability weights would be region—or country—specific, to ensure that the “burden”

measured describes the actual suffering of the people as well as possible. The metric for funding, on the other hand, should be based on economic measures such as GDP, as well as other measures of the resources available in the health system of a country. This two-part system would appease both social scientists and quantitative researchers, and could lead to more equitable spending on global health that would also be more consistent with the needs of present and future populations.

References

1. Anand, S., & Hanson, K. (1997). “Disability Adjusted Life Years: A Critical Perspective.” *Journal of Health Economics* 16:685-702.
2. Collins, F.S. (2010). “Research Agenda: opportunities for research and the NIH.” *Science* 327:36-37.
3. Murray, C. (1994). “Quantifying the burden of disease: the technical basis for disability-adjusted life years.” *Bulletin of the World Health Organization* 72(3): 429-445.
4. Hammer, J., & Berman, P. (1995). “Ends and Means in Public Health Policy in Developing Countries.” In *Health Sector Reform in Developing Countries: Making Health Development Sustainable*, edited by Peter Berman, 37-58. Boston: Harvard University Press.
5. Hammer, J., personal communication, 2013.
6. Reich, M. (1995). “The Politics of Health Sector Reform in Developing Countries: Three Cases of Pharmaceutical Policy.” In *Health Sector Reform in Developing Countries: Making Health Development Sustainable*, edited by Peter Berman, 37-58. Boston: Harvard University Press.
7. Becker, A., Motgi, A., Weigel, J., Raviola, G., Keshavjee, S., & Kleinman, A. (2013). “The Unique Challenges of Mental Health and MDRTB: Critical Perspectives on Metrics of Disease Burden.” In *Reimagining Global Health: An Introduction*, edited by Paul Farmer, Jim Yong Kim, Arthur Kleinman, and Matthew Basilio, 209-241. Berkeley: University of California Press.
8. Arensen, T., & Nord, E. (1999). “The value of DALY life: problems with ethics and validity of disability adjusted life years.” *BMJ* 319: 1423-1425.
9. Reidpath, D., Allotey, P., Kouame, A., & Cummins, R. (2003). “Measuring health in a vacuum: examining the disability weight of the DALY.” *Health Policy and Planning* 18(4): 351-356.
10. The Lancet. (N.d). “Special Issues: Global Health Metrics & Evaluation: Data, Debates, Directions.” Retrieved from: <http://www.thelancet.com/journals/lancet/specialissue>.
11. Poenaru, D., Pemberton, J., Frankfurter, C., & Cameron, B. (2013). “Establishing disability weights for congenital paediatric surgical disease: a cross-sectional, multi-modal study.” *The Lancet* 381:S115.
12. World Health Organization. (2004). *The global burden of disease: 2004 update*. Geneva: WHO.
13. Africa Inland Mission. (N.d). “The Good Life.” Last accessed January 13, 2014. <http://www.aimint.org/can/en/see/stories/107-the-good-life>.
14. Hammer, J. “Adding Apples and Oranges (The Burden of Disease and Public Policy).” Presentation at Princeton University, October 7, 2013.
15. Cooper, R., Osotimehin, B., Kaufman, J.S., & Forrester, T. (1998). “Disease burden in sub-Saharan Africa: what should we conclude in the absence of data?” *Lancet* 351:208-210.
16. Kleinman, A. (1995). *Writing at the Margin*. Berkeley: University of California Press.
17. Horstmann, F., & Lopez, A. (2013). “Strengthening vital registration and vital statistics: a standards-based toolkit.” *The Lancet* 381: 64.
18. Grand Challenges in Global Health. (N.d). “Challenge 13: Develop Technologies that Permit Quantitative Assessment of Population Health Status.” Retrieved from: <http://www.grandchallenges.org/Measure-HealthStatus/Challenges/PopulationHealth/Pages/default.aspx>.
19. Institute for Health Metrics and Evaluation. (N.d). “Christopher J.L. Murray.” Retrieved from: <http://www.healthmetricsandevaluation.org/about-ihme/team/christopher-ji-murray>.
20. Adams, V. (2013). “Metrics of the Global Sovereign: Numbers and Stories in Global Health.” Paper presented at Global Health Colloquium, Princeton University, Princeton, New Jersey, October 11.