Issues in Evaluating Measures of Health Disparity

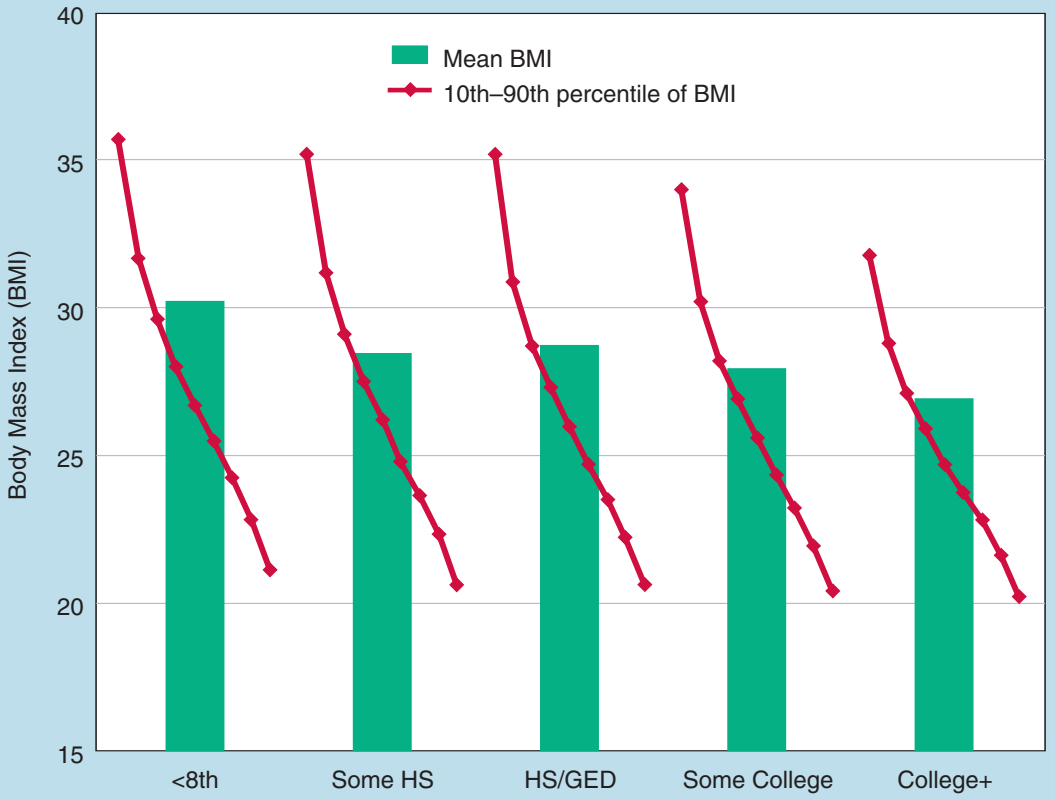
This section discusses several issues—conceptual, pragmatic, and technical—that potentially are important in choosing health disparity measures. Many of these issues receive expanded discussions in the more technical descriptions of the measures that follow in later sections. The intention here is to highlight the set of main issues that might be considered.

Total Disparity vs. Social-Group Disparity

There is an important conceptual issue regarding the specific quantity to be determined when evaluating health disparities. The fundamental distinction to be made is between measuring total disparity, or total variation, and measuring disparities between social groups. The former involves evaluating the univariate distribution of health among all individuals in a population, without regard to their group membership; the latter involves assessing health differences between individuals from certain a priori chosen social groups. The World Health Organization (WHO) initiative to measure health inequality, led by Chris Murray and colleagues, has advocated strongly for an approach to the measurement of health disparity as total health disparity among individuals that is blind to social groups (49,50). Initially, this seems at odds with our notions of why we are evaluating disparity in the first place (51). That is, the initiative to eliminate health disparities arose within the United States because of the persistent presence of social-group health disparities, not out of concern for a widening overall distribution of health. Yet, a deeper understanding of the overall task of determining variation in population health requires that we appreciate the concept of total health disparity. It is likely that the between-group disparity we seek to measure in regard to initiatives such as those in the United States may be relatively small compared to the total disparity that exists between individuals in a population.

Figure 3 (page 20) shows the average body mass index (BMI) for five education groups in the 1997 National Health Interview Survey (NHIS). It is clear that there is a gradient of decreasing BMI with increasing education when comparing average BMI among education groups. The plots of the 10th through the 90th percentiles of BMI, however, show that there is much greater variation in BMI *within* education groups than *between* education groups. Thus, basing the measure of health disparity on between-group average differences may not capture much of the total health variation among individuals. This is not a problematic statement itself but should be understood—and is why indicators of total health inequality can be informative. Thus, based on the group averages and a desire to reduce obesity in the population, focusing a health intervention on the "high-risk" social group (those with less than an 8th-grade education) will in practice target only a limited proportion of those at high risk, because

Figure 3. Mean and 10th–90th Percentiles of Body Mass Index by Education, NHIS, 1997



high-risk individuals exist in every education group.

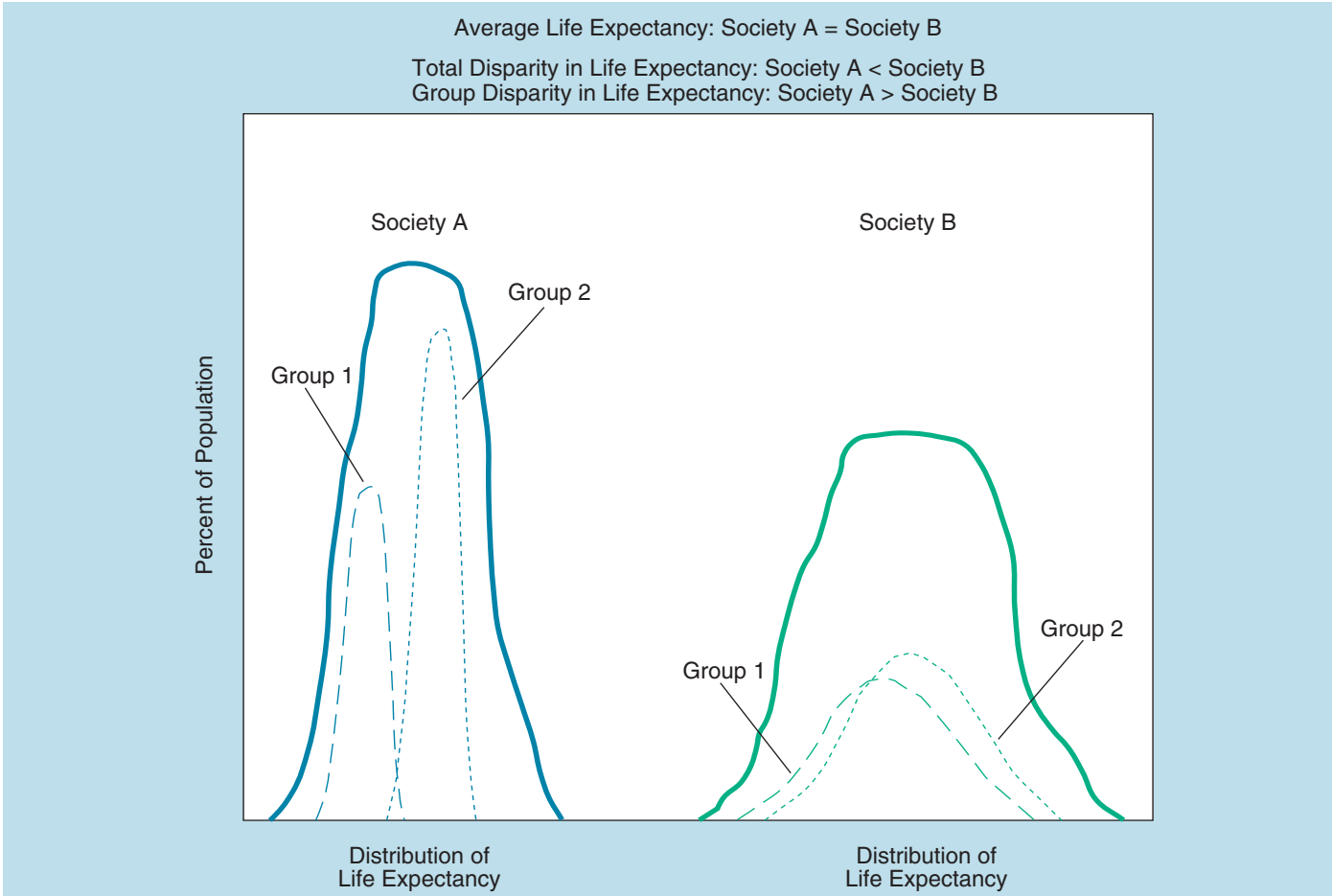
Measures of total health disparity may mask substantial social-group disparities, however. Figure 4 (page 21), adapted from Asada and Hedemann (52), shows the population distributions of life expectancy in two hypothetical societies, A and B. Both populations have the same average life expectancy, but Society A has a much narrower overall distribution of life expectancy; were we to use a measure of total disparity, we would judge Society A to have the smaller disparity. Within Society A, however, there is a substantial gap in life expectancy between social groups 1 and 2, whereas in Society B, groups 1 and 2 have nearly identical life expectancy distributions. If we use a measure of social-group disparity, we likely would judge Society A as having the greater disparity because the distribution of life expectancy between the groups is unequal. The point of this example is to show that measures of total disparity and measures of group disparity may or may not lead to similar judgments about the extent of disparity in two populations or at two time periods. Thus far, the evidence seems to indicate that total disparity and social-group disparity measure different aspects of population health. Two cross-national studies found little correspondence between measures of total disparity and measures of socioeconomic disparity for either child (53) or adult (54) mortality. That is, countries with the largest amount of overall mortality variation did

not necessarily have larger socioeconomic mortality variation, and countries with the largest socioeconomic mortality disparities did not have the largest overall mortality disparities.

Relative and Absolute Disparities

The most frequent method of communicating information about social disparities in public health and epidemiology is in relative terms—through measures of association such as the relative risk. In epidemiology, relative risks are the most common measures of "effect size," partly because they have advantageous properties not shared by absolute risk differences (12,55). Relative and absolute health differences between social groups are the primary language of health disparities, but they provide fundamentally different types of information. Figure 5 (page 22) demonstrates this essential point by showing trends in absolute and relative disparity between males and females in stomach cancer mortality over the past 70 years. Clearly, there was enormous progress in reducing stomach cancer mortality rates among both males and females during the 20th century. As the rates for both groups declined, however, the ratio of male-to-female mortality (i.e., the relative disparity)

Figure 4. Hypothetical Distributions of Life Expectancy in Two Populations

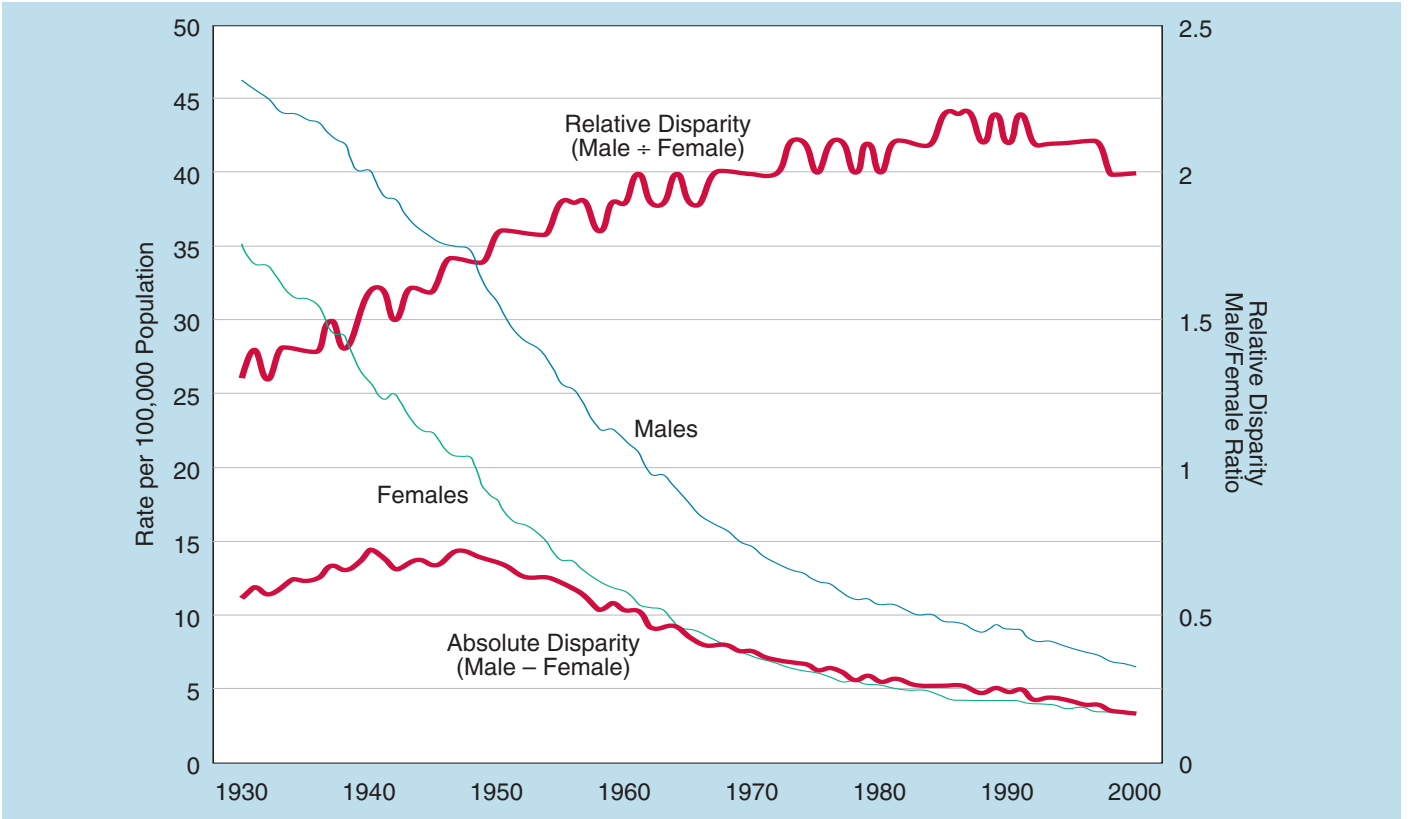


steadily increased. If the difference between male and female mortality (i.e., the absolute disparity) is used as the measure of disparity, however, we observe a different trend. The male-female gap increased from 1930 to about 1950, as female rates declined faster than male rates, and has declined steadily since 1950. Thus, Figure 5 illustrates the possibility that one might arrive at opposite conclusions about what happened to this health disparity, depending on which measure was chosen—the absolute or relative disparity. The reason is that the relative disparity cannot reflect changes in absolute rates—the disparity is relative to the rate in the comparison group.

Reference Groups

The language of disparity—defined literally as "difference"—implies a comparison group. A major question in choosing disparity measures is the choice of comparison group. As noted above, the different definitions of disparity imply different comparison groups, and thus the answer one would get about the extent and patterning of disparity may differ according to which groups are compared. Figure 6 (page 23) shows the situation for cervical cancer mortality rates among several racial/ethnic groups. Hispanic women clearly have the highest incidence of cervical cancer, but how

Figure 5. Absolute and Relative Gender Disparity in Stomach Cancer Mortality, 1930–2000



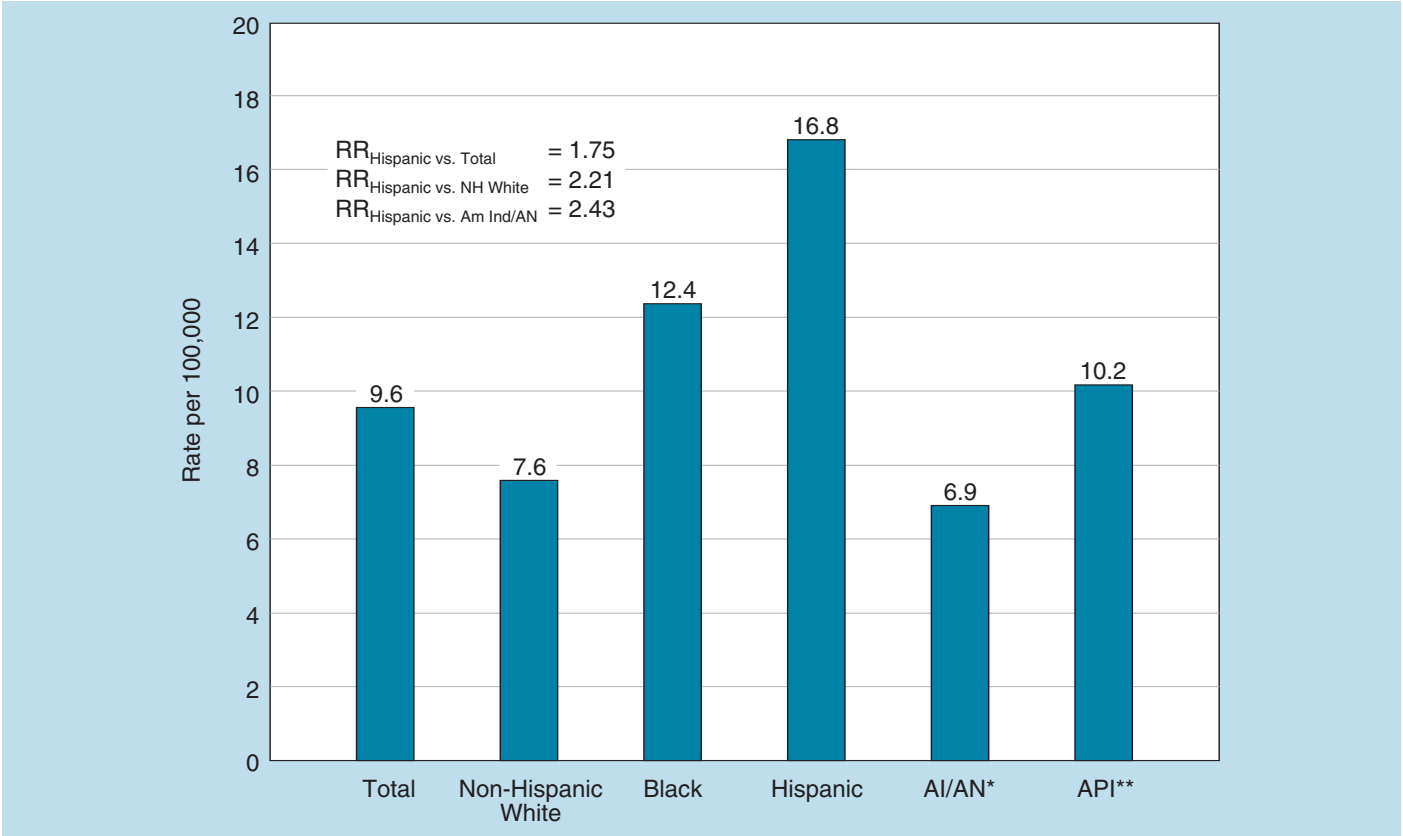
Source: Wingo et al. *Cancer* 2003;97(11 Suppl):3133–275, and SEER Cancer Statistics Review, 1975–2000.

large is the disparity in cervical cancer incidence? The answer depends on the choice of the reference group. If the Hispanic disparity is measured relative to the general population (i.e., the total rate), then the relative disparity is 1.75. If, however, we follow Dr. Satcher's recommendation (48) and focus on the disparity from the majority population—non-Hispanic whites—the relative disparity is 2.21. Or, if the "best-off" group—American Indian/Alaska Natives—is chosen as the reference group, we obtain a relative disparity of 2.43.

Average Population Member

One logical reference group might be the population average, where the disparity measure reflects the gap between the health of different social groups and the mean health of the entire population. The population average is appealing intuitively as a reference point and, as noted above, often is also used explicitly in defining what constitutes a health disparity.

Figure 6. Relative Risk (RR) of Incident Cervical Cancer Among Hispanics According to Varying Reference Groups, 1996–2000



*AI/AN = American Indian/Alaska Native
**API = Asian/Pacific Islander
Source: SEER Cancer Statistics Review, 1975–2000.

Best-Off Group/Person/Rate

This perspective suggests that one might measure disparity as a difference between each social group compared to the healthiest group (or even the healthiest person). This is similar to Sen's concept of shortfalls (56), in which it is assumed implicitly that every social group in the society has the potential to achieve the health of the best-off group. It should be noted, however, that the best-off social group may be relatively small in size, which may lead to substantial variation and instability and could make assessing trends in disparities more difficult.

All Those Better Off

It also is possible to measure disparities by comparison to all those individuals or groups that are better off than a particular group or person. This may seem similar to the "best-off group" reference point, but it differs in a subtle way that may best be illustrated with an example using actual cancer data. Figure 7 (page 25) shows cancer incidence from 1996–2000 by race and ethnicity for two different cancers, kidney/renal pelvis and myeloma. In both cases, there is a substantial difference between the group with the highest incidence rate, blacks, and the group with the lowest or "best" rate, Asian/Pacific Islanders. When we look at the incidence rates of other groups, however, we see two different situations. In the case of kidney cancer, Hispanics and whites have rates more similar to blacks, whereas, in the case of myeloma, they have rates more similar to Asian/Pacific Islanders. Relative to all those better off than blacks, most people might judge the disparity to be worse in the case of myeloma compared to kidney cancer; yet, if measured relative to the "best-off" group perspective, we would be unable to capture this nuance.

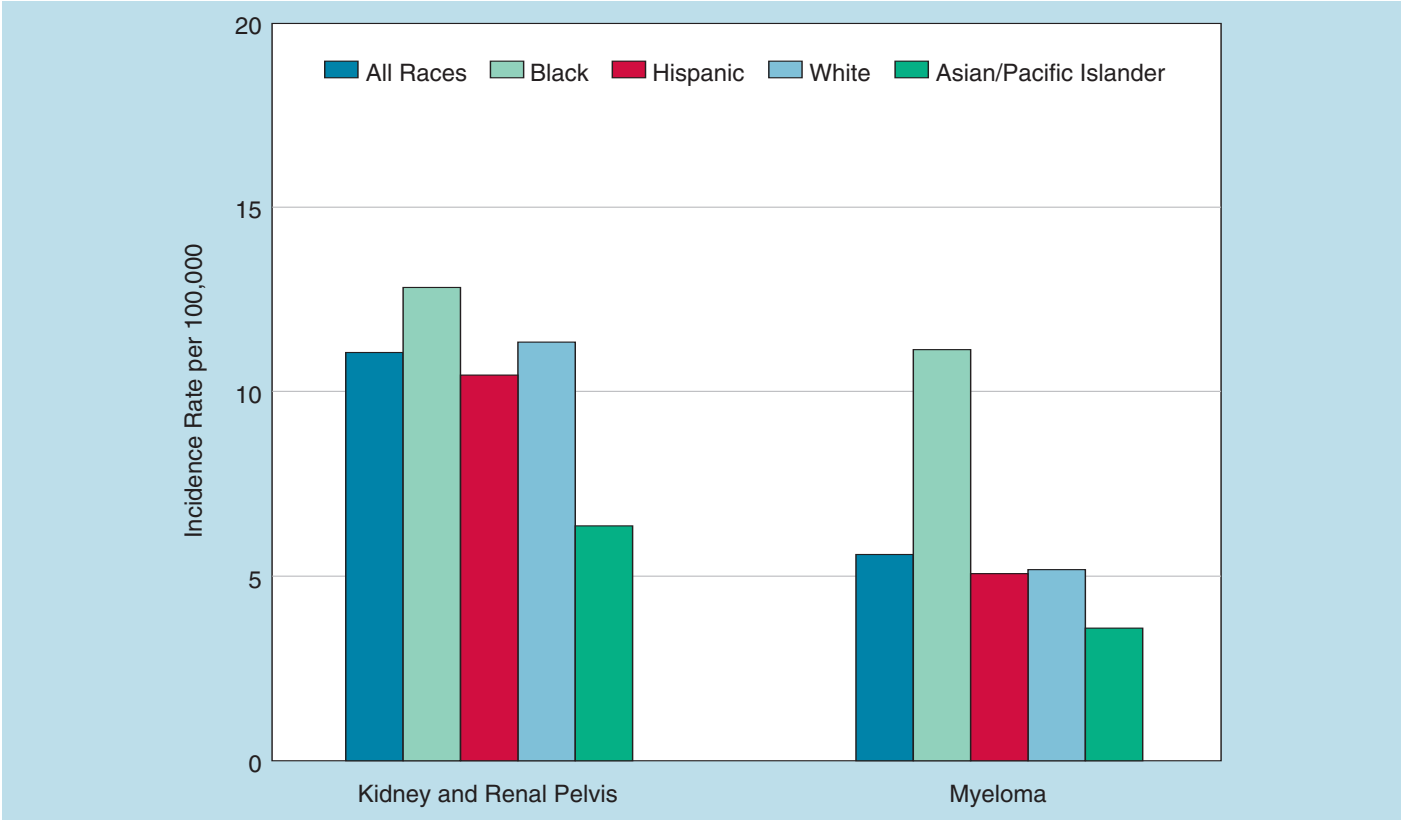
Fixed/Target Rate

The prior three reference groups are inherently relative as they change over time, which may make assessments of trends in disparities inconclusive if using pairwise comparisons. One advantage of a fixed or target rate is that the reference level does not change over time unless a new target is adopted.

Social Groups and "Natural" Ordering

The *Healthy People 2010* initiative mandates eliminating health disparities within a number of different types of social groupings: gender, income and education, disability, geographic location, sexual orientation, and race and ethnicity. Such groupings were chosen because they represent important normative dimensions of U.S. society, and it has been shown repeatedly that health differences exist between these social groups. The above groups, however, also differ in ways that may have implications for monitoring health disparities. The social groups that measure dimensions of socioeconomic position—education and income—have an inherent ordering *regardless of the health status of their members*. Individuals with less than a high-school education unambiguously have less formal education than do individuals with a college degree. The same cannot be said for the other groups targeted by the *Healthy People 2010* initiative. There simply is

Figure 7. Age-Adjusted Incidence of Kidney/Renal Pelvis Cancer and Myeloma by Race and Ethnicity, 1996–2000



Source: SEER Cancer Statistics Review, 1975–2000.

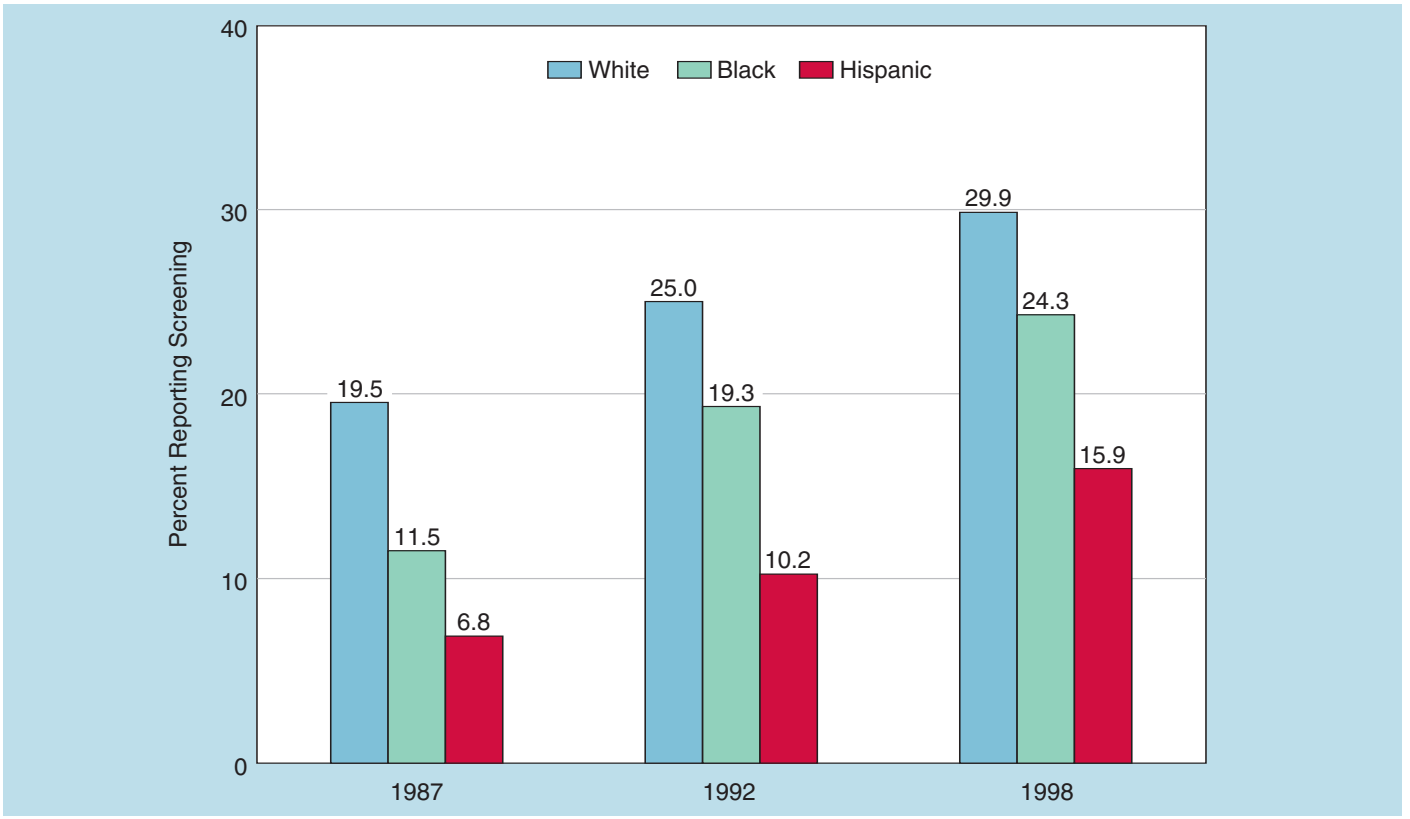
no inherent way to rank individuals by their race, ethnicity, disability status, or sexual orientation. Certain measures of disparity cannot be used to measure or monitor disparities between groups that have no implicit ranking. For example, the slope index of inequality and the concentration and achievement indices cannot be used except in the case of education and income, because there is no inherent way to rank some social groups such as racial/ethnic groups or genders (except by their health level). In the *Healthy People 2010* parlance, groups with a "natural" ordering include education and income but do not include gender, race/ethnicity, sexual orientation, disability status, and geography.

The Number of Social Groups

Should the measure of disparity include information from all social groups (i.e., the entire population), or is it sufficient to reflect only the experiences of the best- and worst-off (extreme) groups? Many empirical studies of health disparities measure disparity by comparing the extreme groups (e.g., the lowest income group compared with the highest income group). This, however, ignores the health status of other groups and additionally may only reflect the disparity between two very small population groups. For example, in 2000 there was a three-fold relative difference in death rates from melanoma of the

skin across U.S. states. The states with the lowest (North Dakota, 1.3 per 100,000) and the highest (Wyoming, 3.7 per 100,000) rates, however, collectively accounted for only 0.4% of the U.S. population in that year. Eliminating this disparity would have little impact on reducing the population burden of melanoma mortality because only a fraction of melanoma cases reside in these two states. Additionally, although there are good reasons for focusing attention on specific comparisons, such as the disparity between blacks and whites in the receipt of treatment for cancers of similar stage (57), such pairwise comparisons do not quantify the disparity across *all* racial/ethnic groups, which is precisely the goal of initiatives to eliminate health disparities by the year 2010. For example, the gap between white and black men in the recent use of fecal occult blood test (FOBT) screening for colorectal cancer narrowed between 1987 and 1998 (58); however, this pairwise comparison conceals the fact that the gap between Hispanics and whites and between Hispanics and blacks increased (see Figure 8). Despite the utility of measuring disparities between two groups, pairwise comparisons may conceal important heterogeneity and thus provide a limited view in monitoring progress toward eliminating health disparities across the entire range of social groups.

Figure 8. Proportion of Men Reporting Use of Screening Fecal Occult Blood Tests (FOBT), by Race and Ethnicity, 1987–1998



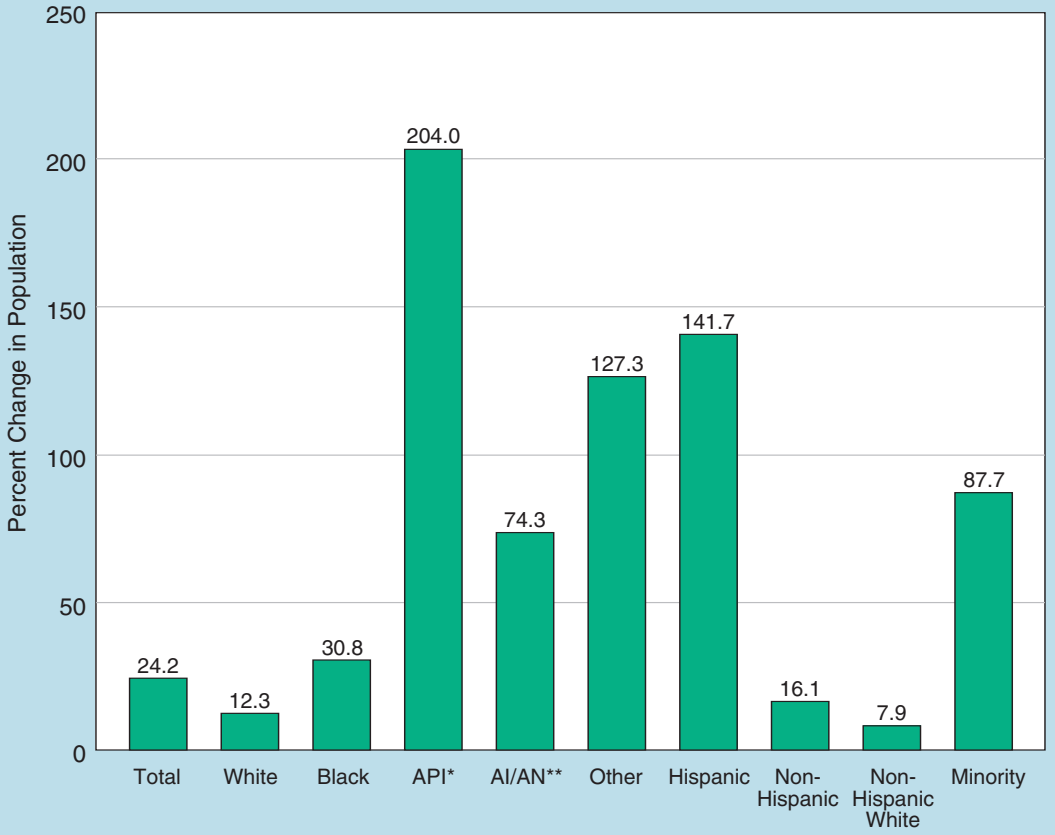
Source: Breen et al. *J Natl Cancer Inst* 2001;93:1704–13.

Population Size

Should the disparity measure incorporate the size of the groups being compared? If we use a pairwise comparison of extreme groups, would it matter that one or both of those groups comprises a very small proportion of the population? For example, Pearcy and Keppel's Index of Disparity (30) gives equal weight to each group, even though the groups may represent different proportions of the population. This has important implications for monitoring disparities and is another case in which a statistical choice reflects an ethical choice. That is, the decision of whether or not to weight social *groups* by their population size also is a decision regarding how much weight to give *individuals* within each social group. For example, if we measure the disparity in prostate cancer mortality among U.S. states in 2000 without weighting states by their population size, California and Wyoming receive equal weight despite the fact that California has nearly 70 times as many males as Wyoming. Thus, in an unweighted analysis of U.S. states, *individual* males in California receive approximately 1/70th the weight of males in Wyoming.

Another important issue in using unweighted measures of health disparity is their inability to incorporate the demographic changes that inevitably occur over time. For example, Figure 9 (page 28) shows the percentage increase in population subgroups between the 1980 and 2000 Census (59). These demographic shifts can have enormous impact on the population's health and should be factored into the assessment of health disparity. In their analysis of the effects of education on all-cause and cause-specific mortality in the American Cancer Society's Cancer Prevention Study cohorts (CPS-I and CPS-II), Steenland and colleagues (who used ordinary least squares regression) noted that changes in the distribution of education made it difficult to compare the extent of disparity between the two populations studied (36). The proportion of the population with less than a high-school education was 20% in CPS-I and 6% in CPS-II, while those with a college degree were 16% and 30% of the population in the two respective cohorts. In epidemiological language, the proportion of the population "exposed" changed dramatically with large population shifts out of the most disadvantaged groups. For a measure of health disparity to allow for an unambiguous comparison across time, it should be sensitive to changes in the distribution of social groups over time. This sensitivity to changes in the proportion of people exposed to disadvantageous social positions especially is important when considering the so-called "upstream" determinants of health disparities. It is commonplace in health disparity research to discuss how distal social policy affects health and health disparity. The policies and programs that define the nature of stratification in a society create educational opportunity, allocate income, and affect the types of jobs that are available. When these "upstream" social policy factors affect the nature of social stratification by reducing the number of minimally educated individuals, for instance, thus reducing the number of individuals exposed to that form of social disadvantage, then measures of health disparity should account for that change. The same situation exists when the proportion of a particular population subgroup changes over time, as in the case of the migration of Hispanics and Asian/Pacific Islanders as shown in Figure 9.

Figure 9. Percent Change in Population Size by Race and Hispanic Origin, 1980–2000



*Asian/Pacific Islander
**American Indian/Alaska Native
Source: Hobbs F, Stoops N. *Demographic Trends of the 20th Century,* 2002.

Socioeconomic Dimension

Another potential criterion for a measure of health disparity, first articulated by Wagstaff and colleagues (60), is whether the measure is able to capture health gradients associated with socioeconomic position. By health gradients, we mean a situation where a measure of health status either increases or decreases with increasing socioeconomic position. A good example is the increasing rate of cancer incidence among individuals living in U.S. counties with successively higher poverty rates (32). That is, is the measure sensitive to the *direction* of the association between social group and health? For instance, if at one time health status increases with social-group ordering and at another time health decreases with the same social-group ordering, the disparity measure will reflect this change if it is sensitive to the direction of the gradient. Of necessity, this criterion is applicable only for measuring inequality between social groups that have an inherent ranking. The lack of inherent ordering among racial/ethnic groups, for example, means that the "socioeconomic dimension" criterion cannot be applied to disparity measures used to monitor racial/ethnic health disparities.

Monitoring Over Time

Inherent in the goals of *Healthy People 2010* is the idea that we monitor progress toward the elimination of health disparities. That means it is desirable that measures of disparity are interpretable over time. This represents important challenges for the use of simple pairwise relative disparity indicators and indicators that are not population-weighted. Because *both* the health status within different social groups and the population distribution of social groups change over time, which together reflects the overall public health burden of health disparities, measures that are sensitive to both dimensions of change may be more suitable for monitoring disparities over time.

Transfers

The issue of how measures of disparity respond to hypothetical transfers between individuals has been an important part of evaluating the performance of income disparity measures in economics. The major test in economics is the principle of transfers—sometimes called the Pigou-Dalton condition (61,62)—which maintains that a transfer of income from a richer to a poorer person should result in a decrease in the measure of disparity (assuming that everyone else's income remains unchanged and the transfer is not large enough to reverse anyone's relative positions). This is an intuitively powerful and desirable notion that corresponds well with what we believe disparity measures should be able to capture. Yet, theoretically, this is a somewhat difficult concept to employ for judgments about health disparity. Is "health" a fungible good like income that can be redistributed in different ways? It is hard to imagine social mechanisms (perhaps apart from organ donation) through which a "healthy" person can directly transfer some of her health to someone who is less healthy, though it is possible to conceive of redistributing health resources. The task, however, is to measure disparities in health, not health resources.

We have noted that measuring disparity in health versus income differs in at least one important respect, namely that goods such as income or wealth are, in fact, transferable from one individual to another. One potential way to avoid this difficulty is to think of comparing disparity in two different populations (e.g., in two repeated observations of a cohort). One might then think of a transfer-like principle according to which we evaluate a measure of health disparity. If the health of every individual remains the same, but a single "healthier" person becomes less healthy and a previously "less healthy" person's health improves, the measure of health disparity should decrease (25). This seems a plausible-enough principle to warrant evaluating a measure of health disparity, but health disparities and income distributions are dissimilar in another way. Even if we are willing to put aside the issue of the literal inability to "transfer" health, it is not at all clear in the previous example that we would be willing to accept the decreased health of one person for the sake of increasing the health of another. For income, this is not a problem because it is the distribution of the good itself that is under question. Most people generally believe that it is unfair that some have enormous incomes while others live in extreme poverty. Do

we truly believe, however, that some individuals possess *more* than their fair share of health? As was emphasized earlier, one of the major reasons for the increasing focus on health inequalities is not simply that some are healthy while others are sick. It is that some *kinds* of individuals or the members of some social groups are healthy while other kinds are sick. It is the normative distinction between the kinds of healthy or unhealthy individuals that drives our concern that health differs so markedly by social group. The concern over health disparities then, at least in the current historical period, is not that there are health differences in society but that these health differences systematically covary with membership in particular social groups.

Subgroup Consistency

Generally, this criterion says that if the measure of overall disparity includes, for example, three groups, and disparity within two groups remains unchanged while increasing within the third, the measure of disparity should increase. This is of most relevance when we are interested in measuring the overall disparity (i.e., across the entire population). For instance, suppose we were examining alcohol consumption at two points in time in a population composed of two social groups (rich and poor). At each time, both the size of these groups and their average alcohol consumption remain constant, but the disparity in consumption increases within the poor and remains constant within the rich. Subgroup consistency requires that any measure of overall disparity also should register an increase in this scenario. This is not likely to be an important criterion for health disparity measures in the context of *Healthy People 2010* because it does not focus on health disparities within subgroups of a social group (e.g., within the poor).

Decomposability

Decomposition as a property of statistical measures is common in both economics and epidemiology. In economics, it typically refers to the ability to decompose a measure of disparity by sources of income or into between-group and within-group partitions (40). Decomposable disparity measures are seen as advantageous as they can offer information about the sources of increasing or decreasing disparity as indicated in a summary statistic. In public health, decomposition often is used to capture differences in summary rates. For example, a difference in age-adjusted mortality rates between two populations can be "decomposed" into differences between mortality rates and differences in age structure.

Scale Independence

Scale independence (or invariance) often is seen as a desirable property of disparity measures. It often is argued that, all else being equal, if everyone's health "doubles," the disparity measure should remain unchanged. It is arguable, however, whether for public health, where we also are concerned about the absolute level of ill health, this is a desirable property.

Transparency/Interpretability for Policy Makers

Finally, it seems salient that an interpretability criterion be included as a factor in decisions about measures of disparity. For instance, despite other desirable properties, the actual value of more sophisticated summary measures such as the Concentration Index have no obvious interpretation and thus may make communicating health disparity indices to the community and policy makers potentially more difficult. Thus, the extent to which different measures of disparity can be captured graphically to aid communication might be important in deciding which measures are most appropriate for monitoring cancer-related health disparities. Perhaps the use of real-time graphical displays of changes in outcomes of interest may aid the understanding of health disparity. This dimension of health disparity monitoring should not be underestimated, as evidenced by the lack of general application of more sophisticated disparity measures in health disparity research.