Appendix: Random Variation

While this report did not focus on statistical inference about changes in health disparities, it is likely to be of interest to many researchers and policymakers. Therefore, in this Appendix we provide a very basic introduction to the various methods used for calculating standard errors for the summary measures of disparity discussed in this report. We would encourage those interested to consult the source publications referenced here for more details.

Underlying Rates

Most of the underlying data in this report are based on either incidence or mortality rates and are assumed to come from a Poisson distribution (17). The general formula for the standard error for crude or age-specific mortality and incidence rates is:

$$SE_r = \frac{\sqrt{d}}{n} \qquad [12]$$

where SE is standard error, d is the number of incident cases or deaths, and n is the estimated population size (17).

Rate Ratio and Rate Difference

The general formula for calculating the standard error of the Rate Difference (RD) is (12):

$$SE_{RD} = \frac{r_1 - r_0}{\sqrt{SE_{r1}^2 + SE_{r0}^2}} \qquad [13]$$

where r_0 is the reference rate and SE indicates the standard error of the rates being compared. Similarly, the standard error for the rate ratio, assuming the two estimates are independent, may be written as (18):

$$SE_{RR} = \frac{r_1}{r_0}\sqrt{\left(\frac{SE_{r1}}{r_1}\right)^2 + \left(\frac{SE_{r0}}{r_0}\right)^2} \qquad [14]$$

though it is often more convenient to work with the natural log of the RR for generating confidence intervals:

$$SE_{RR} = \frac{r_1}{r_0}\sqrt{\left(\frac{SE_{r1}}{r_1}\right)^2 + \left(\frac{SE_{r0}}{r_0}\right)^2} \qquad [15]$$

Slope and Relative Index of Inequality

Recall that the SII can be easily obtained via regression on grouped data:

$$SE_{\log(RR)} = \sqrt{\frac{1}{d_1} - \frac{1}{n_1} + \frac{1}{d_0} - \frac{1}{n_0}} \qquad [16]$$

where j indexes social group, $\bar{y}_j$ is the average health status and $\bar{R}_j$ the average relative ranking of social group j in the cumulative distribution of the population, β_0 is the estimated health status of a hypothetical person at the bottom of the social group hierarchy (i.e., a person whose relative rank R_j in the social group distribution is zero), and β_1 is the difference in average health status between the hypothetical person at the bottom of the social group distribution and the hypothetical person at the top (i.e. R_j=0 vs. R_j=1). However, because this regression is on grouped data, the standard errors are heteroskedastic, and Kakwani and colleagues, and Low and Low (19), among others, note that the following transformation should be used, which is equivalent to running weighted ordinary least squares:

$$\bar{y}_j \sqrt{P_j} = \sqrt{P_j} + \beta_1 \bar{R}_j \sqrt{P_j} \quad [17]$$

where p_j is the proportion of the population in the jth group. However, this transformation does not account for the correlation of error terms induced by the relative ranking variable. See Kakwani and colleagues (8), Wagstaff (20), or Low and Low (19) for additional details.

Hayes and Berry (21) have also developed a formula for the standard error for the Relative Index of Inequality (RII) for grouped data, using the Kunst-Mackenbach version of the RII, which may, via substitution, be re-written as:

$$RII_{KM} = \gamma = \frac{\alpha}{\alpha + \beta} = \frac{\bar{y} - \beta\bar{R}}{\bar{y} - \beta(1 - \bar{R})} \quad [18]$$

in which case the estimated variance and standard error of γ are:

$$\text{var}(\gamma) = \frac{\beta^2 \text{var}(\bar{y}) + \bar{y}^2 \text{var}(\beta)}{[\bar{y} + \beta(1 - \bar{R})]^4}$$

$$SE(\gamma) = \sqrt{\text{var}(\gamma)} \quad [19]$$

Relative Concentration Index

The formula for the standard error for the Relative Concentration Index (RCI) for grouped data is given by Kakwani and colleagues, which accounts for the autocorrelation in error terms, as:

$$\text{var}(RCI) = \frac{1}{n}\left[\sum_{j=1}^{J} p_j a_j^2 - (1 + RCI)^2\right] + \frac{1}{n\mu^2}\sum_{j=1}^{J} p_j \sigma^2_j (2R_j - 1 - RCI)^2 \quad [20]$$

where σ^2 is the variance of μ,

$$a_j = \frac{\mu_j}{\mu}(2R_j - 1 - C) + 2 - q_{j-1} - q_j \quad [21]$$

$$q_j = \frac{1}{\mu}\sum_{\gamma=1}^{j} \mu_\gamma p_\gamma$$

where n is the sample size, J is the number of groups, p_j is the proportion of the total population in group j, μ_j is the mean value of the health variable in group j, and *RCI* is the Relative Concentration Index. A very useful guide for calculating the RCI and its standard error is available at the World Bank's 'Poverty and Health' website (Accessed 30 July 2007) at: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTPAH/0,,contentMDK:20216933~menuPK:400482~pagePK:148956~piPK:216618~theSitePK:400476,00.html

Other Summary Measures

To our knowledge, details of standard formulas for calculating standard errors for other summary measures used in this report, including the Index of Disparity, the Theil Index and Mean Log Deviation, and the Between-Group Variance, are not widely available. Cowell (22, 23) gives an overview for various measures of income inequality, and Biewen and Jenkins (24) recently derived methods for estimating various inequality measures with complex survey data. However, recent developments in computing power and resampling methods (25) have led many authors to suggest calculating standard errors via bootstrapping (12, 26-29), which was the approach we took in Case Study 1.