Methods and Software for Estimating Health Disparities: The Case of Children’s Oral Health

Nancy F. Cheng, Pamela Z. Han, and Stuart A. Gansky

Received for publication December 7, 2007; accepted for publication June 12, 2008.

The National Center for Health Statistics recently issued a monograph with 11 guidelines for reporting health disparities. However, guidelines on confidence intervals (CIs) cannot be readily implemented with the complex sample surveys often used for disease surveillance. In the United States, dental caries (decay) is the most common chronic childhood disease—5 times more common than asthma. Racial/ethnic minorities, immigrants, and persons of lower socioeconomic position (SEP) have a greater prevalence of caries. The authors provide methods for applying National Center for Health Statistics guidelines to complex sample surveys (health disparity indices and absolute and relative difference measures assessing associations of race/ethnicity and SEP to health outcomes with CIs); illustrate the application of those methods to children’s untreated caries; provide relevant software; and report results from a simulation varying prevalence. They use data on untreated caries from the California Oral Health Needs Assessment of Children 2004–2005 and school percentage of participation in free/reduced-price lunch programs to illustrate the methods. Absolute and relative measures, the Slope Index of Inequality, the Relative Index of Inequality (mean and ratio), and the Health Concentration Index were estimated. Taylor series linearization and rescaling bootstrap methods were used to estimate CIs. Oral health differed significantly between White children and all non-White children and was significantly related to SEP.

A health disparity can be defined as occurring when 1 population group carries a disproportionate share of a health burden (1). The National Institutes of Health Strategic Research Plan to Reduce and Ultimately Eliminate Health Disparities defines a health disparity as a “difference in incidence, prevalence, mortality, and burden of disease and other adverse health conditions that exist among specific population groups in the United States” (2, p. 7). The National Center for Health Statistics recently issued a monograph with 11 guidelines for reporting health disparities (3). Although this thorough monograph discusses relevant issues, makes recommendations, and provides illustrative worked examples, not all of the guidelines can be readily implemented, especially with the complex sample surveys that are often used for disease surveillance. For example, estimating standard errors or confidence intervals for health disparity indices such as the Health Concentration Index (3) is not straightforward. Custom software programs are needed for performing recommended analyses.

The Surgeon General’s report on oral health (4) draws attention to the “profound and consequential disparities in the oral health of our citizens” and the “silent epidemic” of dental and oral diseases that affect particular groups. The Public Health Service’s Healthy People 2010 initiative outlines oral health objectives addressing these disease patterns (5). These objectives fall under one of Healthy People 2010’s primary goals: to eliminate health disparities within certain segments of the population, including racial/ethnic minorities.
Early childhood caries is generally considered an infectious tooth in a child 71 months of age or younger (12, p. 13). In the United States, dental caries or tooth decay is the most common chronic disease of childhood—5 times more common than asthma and 7 times more common than hay fever (4). The disease burden of dental caries is disproportionately distributed in US children, with 75% of caries occurring in 8.1% of 2- to 5-year-olds (primary dentition) and 33.0% of children aged 6 years or older (permanent dentition) (6). Polarization of dental caries has resulted in certain groups (racial/ethnic minorities, immigrants, and persons of lower SEP) experiencing a larger burden of disease (7–11).

Dental caries in very young children (age <6 years), known as early childhood caries, is defined as “the presence of 1 or more decayed (noncavitated or cavitated lesions), missing (due to caries), or filled tooth surfaces in any primary tooth in a child 71 months of age or younger” (12, p. 13). Early childhood caries is generally considered an infectious disease, with Streptococcus mutans being its primary bacterial infectious agent upon eruption of teeth in infants (13). In addition to the short-term consequences of enduring pain and toothaches, children with early childhood caries are more likely to continue to have dental problems as they grow older (14).

Oral health has generally improved in the US population during the last decade, except for the prevalence of early childhood caries, which has increased from 24% to 28% (15). Even with the declining rates of dental caries in the US population, oral health disparities continue to be an ongoing problem (15, 16). In 2- to 10-year-olds above the poverty level, indices of caries prevalence showed a decrease of 48% from 1971–1974 to 1988–1994, while indices for those at or below the poverty level showed a decrease of only 23% (16). In addition, during 1988–1994, children from poor families suffered 12 times more restricted-activity days from dental disease than children from higher-income families (4). Among low-income families, 50% of dental caries remained untreated (4).

In this paper, we provide methods for applying National Center for Health Statistics guidelines to complex sample surveys, such as those using health disparity indices and measures of absolute and relative differences, to assess associations of race/ethnicity and SEP to health outcomes with confidence intervals; illustrate the application of those health disparity methods to children’s oral health; provide free, widely available software with which to calculate estimates from health disparity indices; and examine the behavior of health disparity indices as a function of disease prevalence with a simulation.

MATERIALS AND METHODS

Description of data

The California Oral Health Needs Assessment of Children (COHNAC) 2004–2005 was a complex, stratified cluster-sample survey of kindergarteners and third-graders (n = 21,399) from 186 California elementary schools. Strata were geographic regions, and schools were clusters. Sampling weights incorporated both inverse selection probability (school cluster and child) and nonresponse. This survey followed the general Association of State and Territorial Dental Directors’ surveillance model that is used in many states. The State of California institutional review board waived review of the survey, since it was a surveillance activity; the University of California, San Francisco, institutional review board approved additional analyses from the survey as being exempt research. Children from the schools participated in dental screenings which assessed several measures of dental health, including untreated caries. Participation in a free or reduced-price lunch (FRL) program was the key measure of SEP. School-level FRL information was obtained from the 2005 Academic Performance Index database, downloaded from the California Department of Education website. Guidelines described by Keppel et al. (3) were applied to the COHNAC 2004–2005 data as follows. Weighted analysis provides more information about the effect a disparity has on the population as a whole, rather than just within smaller groups. Weighted results are considered primary analyses in this stratified cluster survey. Unweighted results can be downloaded from the University of California, San Francisco, website given in the Acknowledgments.

Absolute and relative measures

Disparities were measured as separation of each group from the most favorable (lowest) prevalence group. In this study, we measured disparities in terms of an adverse event (e.g., disease rather than absence of disease). The reference group for the race/ethnicity category was non-Hispanic Whites. The reference groups for SEP were not being a participant in an FRL program and attending a school with less than 25% FRL program participation. Both relative and absolute measures were used to compare disparities. The absolute disparity, measured as the simple difference, was calculated as the difference between the prevalence for each group and the prevalence for the reference group. The relative disparity, measured as percent difference, was determined as 100 times the simple difference divided by the reference group prevalence. Ninety-five percent confidence intervals were calculated for both absolute and relative measures as

\[ T \pm 1.96 \times \text{SE}_p \]

where \( T \) is the point estimate for a statistic and \( \text{SE}_p \) is the standard error (SE) of \( T \). With mutually exclusive groups, the standard error of the simple difference (diff), \( \text{SE}_{\text{diff}} \), is computed as

\[ \text{SE}_{\text{diff}} = \sqrt{\text{SE}_i^2 + \text{SE}_r^2} \]

where \( i \) is the standard error of a non-reference-group prevalence and \( r \) is the standard error of the reference-group prevalence. With available standard errors, the statistical significance of the simple difference can be tested using a \( Z \) statistic,

\[ Z = (R_i - R_r)/\sqrt{\text{SE}_i^2 + \text{SE}_r^2} \]

where \( R_i \) is the non-reference-group prevalence and \( R_r \) is the reference-group prevalence.

The standard error of the percent difference (pctdiff), \( \text{SE}_{\text{pctdiff}} \), is computed on the basis of the relative standard

Am J Epidemiol 2008;168:906–914
error (RSE) of the percent difference, RSE_{pctdiff}, as \( SE_{pctdiff} = RSE_{pctdiff} \times \text{pctdiff} \), where RSE_{pctdiff} is obtained from the relative standard error for the simple difference (RSE_diff) and the relative standard error for the reference group (RSE_r) as \( RSE_{pctdiff} = \sqrt{RSE_{diff}^2 + RSE_r^2} \). RSE_{diff} is computed as \( RSE_{diff} = \sqrt{SE_{diff}^2 + SE_r^2} / (R_i - R_e) \), and RSE_r is calculated from the standard error as \( RSE_r = SE_r / R_e \). To evaluate the statistical significance of percent difference, the Z statistic is computed as \( \text{pctdiff} / SE_{pctdiff} \). The statistic is computed as \( \text{pctdiff} / SE_{pctdiff} \). The \( P \) values for both the simple difference and percent difference can then be easily calculated from the Z statistic. With multiple categorical groupings such as race/ethnicity, \( Z \) groups (population for each group is computed. The size of each group 

First, SEP groups are ranked from lowest to highest on the form \( y \) against each SEP group's proportion range. Weighted least squares are used with the weights proportional to the population size of each group. The fitted regression line has the form \( y = \alpha + \beta x \). The regression slope \( \beta \) is designated the Slope Index of Inequality (SII).

SII is interpreted as the average difference in percentage of adverse health in the population ranked from the lowest SEP group to the highest. A negative SII indicates that the adverse health indicator decreases with increasing SEP, while a positive value means that the health indicator rises with increasing SEP; 0 indicates no association between adverse health and SEP.

Two versions of the Relative Index of Inequality (RII) are related to the SII. RII(mean) = \( \frac{\beta}{\bar{y}} \), where \( \bar{y} \) is the overall mean prevalence of the weighted group values, and RII(ratio) = \( \frac{\alpha}{\alpha + \beta} \), which measures the ratio of prevalence at the highest SEP level \( y \) at \( x = 0 \) to prevalence at the highest SEP level \( y \) at \( x = 1 \). RII(ratio) tends to amplify extreme differences (very large or very small) in comparison with RII(mean), much like an odds ratio compared with a difference of proportions. Note that SII is sensitive to the mean prevalence rate of the population (19). For example, if adverse health doubled, then mean prevalence and SII would double, but RII(mean) and RII(ratio) would remain the same.

Ninety-five percent confidence intervals are calculated for SII, RII(mean), and RII(ratio) as \( S \pm c \times SE_S \), where \( S \) is the point estimate for SII, RII(mean), or RII(ratio), \( SE_S \) is the standard error for \( S \), and \( c \) is the critical 5% value from a \( t \) distribution with \( g - 2 \) df, with \( g \) being the number of SEP groups.

The standard error for each health disparity index \( S \) is derived as follows. Assuming a linear relation between the prevalence and SEP groups, the variance of SII can be estimated from the standard deviation about the fitted regression line. The variance of RII(mean) = \( var(\beta/\bar{y}) = var(\beta)/\bar{y}^2 \). Hayes and Berry (20) considered the calculation of the standard error for RII(ratio) = \( \frac{\alpha}{\alpha + \beta} = \gamma \) using a first-order Taylor series linearization (21, 22) and recommended working with the natural logarithm of \( \gamma \), which gives us

\[
\text{var}[\ln(\gamma)] = \frac{[\beta^2\text{var}(\bar{y}) + \bar{y}^2\text{var}(\beta)]}{\{(\bar{y} - \beta x)^2(\bar{y} + \beta(1 - x))^2\}}.
\]

The 95% confidence interval for \( \gamma \) is then \( \exp(\ln(\gamma) \pm c \times SE[\ln(\gamma)]) \), where \( \text{SE}[\ln(\gamma)] = \sqrt{\text{var}[\ln(\gamma)]} \) and \( c \) is the critical 5% value from a \( t \) distribution with \( g - 2 \) df, with \( g \) being the number of SEP groups.

An alternative to regression-based methods is the Health Concentration Index \( C \) based on the concentration curve, which is similar to the Gini coefficient from the Lorenz curve. \( C \) plots the cumulative proportion of the population, ranked by SEP level from the lowest SEP category to the highest, along the horizontal axis against the group’s share of adverse health along the vertical axis. It is defined as twice the net area between the concentration curve and the diagonal \( (45^\circ) \) line of equality. Values of \( C \) range from \( -1 \) to \( +1 \). \( C \) is negative when the concentration curve lies above the diagonal, indicating that adverse health declines uniformly with increasing SEP category; positive when the curve lies below the diagonal; and 0 when it lies on the diagonal, indicating that perfect equality exists in adverse health across SEP. \( C \) is computed using the following formula (23):

\[
C = 2A = 2(1/2 - \text{AUC}) = 1 - 2\text{AUC}
\]

\[
= 1 - \sum_{i=0}^{g-1} (x_{i+1} - x_i)(y_{i+1} + y_i),
\]

where \( A \) denotes the net area between the concentration curve and the diagonal, AUC denotes the area under the concentration curve, \( y_i \) is the cumulative proportion of adverse health in the \( i \)th SEP group, \( x_i \) is the cumulative proportion of the population in the \( i \)th SEP group, and \( g \) is the number of SEP groups. Note that \( x_0 = 0 \) and \( y_0 = 0 \).

To estimate the variance of \( C \), a first-order Taylor series linearization is used. The 95% confidence interval for \( C \) is then computed. The rescaling bootstrap method of Rao et al. (24) for calculating confidence intervals in complex sample surveys is used for comparison. Rather than performing true bootstrap resamples, this method is much more computationally efficient by merely changing the sampling weights for each resample. Details are given in the Appendix.
PROC SURVEYMEANS (SAS, version 9; SAS Institute Inc., Cary, North Carolina), accounting for strata and clusters, was used to estimate prevalences and their standard errors. PROC CROSSTAB (SUDAAN 9.0.1; Research Triangle Institute, Research Triangle Park, North Carolina) was employed in combination with PROC IML to estimate the covariance matrix used to calculate the standard error of \( C \). PROC REG was used to calculate SII and its confidence interval. We developed SAS macros incorporating these steps for calculating simple difference, percent difference, RII(mean), RII(ratio), \( C \), and their standard errors and confidence intervals. Free, publicly available software (SAS macros) for conducting these analyses can be downloaded from a University of California, San Francisco, website (see Acknowledgments).

Simulation study

We undertook a small simulation study to examine the behavior of the absolute and relative measures as well as health disparity indices as the prevalence rate of disease changed. Using the COHNAC survey data, we simulated a binary outcome measure to have an overall prevalence of disease ranging from 10% to 90%. The probability was simulated to have the same linear relation to a 4-category ordinal SEP measure as untreated caries did to school percentage of FRL program participation in the COHNAC survey. The simulated logistic probability model was

\[
P = \frac{\exp\{-0.5 + (-0.3064)\text{SEP} + \varepsilon\}}{1 + \exp\{-0.5 + (-0.3064)\text{SEP} + \varepsilon\}}
\]

with randomly generated \( \varepsilon \); 9 cutpoints for the probability were used to provide overall prevalences of a binary outcome from 0.1 to 0.9. This simple simulation preserved the complex sample survey structure (strata, clusters, and sampling weights) and used the same distribution of the SEP measure. The health disparity indices and 95% confidence intervals were then estimated for the simulated data; the rescaled bootstrap method used 500 bootstrap resamples.

RESULTS

Absolute and relative measures

Table 1 presents absolute and relative measures of weighted prevalence deviations for untreated caries in California Kindergarteners and Third-Graders, 2004–2005.
White reference group. Similarly, there were statistically significant differences between FRL program participants and nonparticipants. Levels of school percentage of FRL participation also differed significantly from the reference group (<25% FRL participation).

Indices of health disparities

Table 2 shows the estimates from health disparity indices of untreated caries relative to SEP, measured as the percentage of FRL participation at schools. SII was estimated as −25.8 (95% confidence interval: −29.5, −22.0), meaning that the prevalence of untreated caries decreased an average of 26% over the population from the lowest SEP to the highest SEP (school percentage of FRL participation). The RII(mean) of −0.91 indicated that the SII was 0.91 times the overall prevalence (28.3%). The RII(ratio) of 2.67 (95% confidence interval: 2.15, 3.32) implied that the regression-predicted prevalence of untreated caries for children from schools with the highest percentage of FRL participation (school with the lowest SEP) was more than 2.6 times the predicted prevalence in schools with the lowest percentage of FRL participation (highest SEP); confidence intervals excluding 1 indicated that the prevalence of untreated caries was significantly related to SEP.

Figure 1 illustrates the relation between school percentage of FRL participation and untreated caries prevalence, while Figure 2 shows the health concentration curve. Since the concentration curve is completely above the line of equality, untreated caries declines uniformly with increasing SEP levels (school levels of percentage of FRL participation); confidence intervals for the Health Concentration Index excluding 0 indicated that the relation between SEP and untreated caries is significant. The rescaled bootstrap confidence interval estimation with 500 bootstrap resamples had narrower intervals than the first-order Taylor series linearization (Figure 5).

DISCUSSION

In this paper, we present practical ways to comply with National Center for Health Statistics guidelines (3) on

<table>
<thead>
<tr>
<th>Health Disparity Index</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope Index of Inequality</td>
<td>−25.8</td>
<td>−29.5, −22.0</td>
</tr>
<tr>
<td>Relative Index of Inequality for the mean</td>
<td>−0.91</td>
<td>−1.04, −0.78</td>
</tr>
<tr>
<td>Relative Index of Inequality for the ratio</td>
<td>2.67</td>
<td>2.15, 3.32</td>
</tr>
<tr>
<td>C (first-order Taylor series linearization)</td>
<td>−0.14</td>
<td>−0.20, −0.07</td>
</tr>
<tr>
<td>C (rescaled bootstrap)</td>
<td>−0.14</td>
<td>−0.16, −0.11</td>
</tr>
</tbody>
</table>

a 95% confidence interval excludes 0.
b Health Concentration Index.
c Rescaled bootstrap 95% confidence interval with 500 resamples (B = 500).

Figure 1. Percentage of California kindergarteners and third-graders with untreated caries in 2004–2005, by school percentage of participation in free/reduced-price lunch programs (%FRL) (≥75%, 50–<75%, 25–<50%, or <25%). Linear regression equation: $y = 41.2 – 25.8x$. Figure 2. Health concentration curve for untreated caries in California kindergarteners and third-graders in 2004–2005, plotted according to school percentage of participation in free/reduced-price lunch programs (%FRL).
reporting health disparities, with oral health disparities used as an illustration. In particular, we provide information about estimating confidence intervals for health disparity indices in complex sample surveys often used for public health surveillance. With a simple simulation, we illustrate how prevalence can affect some measures—notably the relative health disparity measure (percent difference), SII, and C. We also provide free, widely available software (SAS macro) for estimating indices and their confidence intervals.

Moreover, we illustrate that untreated caries (tooth decay) in California kindergarteners and third-graders is a function of SEP, as measured by school percentage of FRL.

Figure 3. Simulation illustrating the effect of disease prevalence (proportion) on absolute and relative health disparity measures for ≥75% school participation in free/reduced-price lunch programs (≥75% FRL). Bars, 95% confidence interval.

Figure 4. Simulation illustrating the effect of disease prevalence (proportion) on 3 health disparity indices (SII, RII(mean), and RII(ratio)), according to school percentage of participation in free/reduced-price lunch programs (≥75%, 50–<75%, 25–<50%, or <25%). RII(ratio), Relative Index of Inequality for the ratio; RII(mean), Relative Index of Inequality for the mean; SII, Slope Index of Inequality. Bars, 95% confidence interval.
participation. This finding is similar to others reported in the literature (4, 7–11), but to our knowledge this is the first report using health disparity indices rather than mere significance testing of prevalences. Caries is the most common chronic disease in children (4), and dental care is the greatest unmet health-care need in US children (25). Since caries is strongly related to school percentage of FRL participation, public health screenings and interventions in schools with high percentages of FRL participation may be a cost-effective targeting method.

Absolute and relative (percent difference) measures of health disparities provide consistent results with the use of cross-sectional data. However, when comparing health disparities among times, regions, or other groups, health disparity indices can provide different interpretations, since, for example, a group with health improving over time but not as much as that of the reference group will have a decrease in the absolute disparity measure but can have an increase in the relative (percent difference) disparity measure. Thus, guideline 3 in the National Center for Health Statistics monograph is to calculate both absolute and relative measures to obtain the full picture. (See the National Center for Health Statistics monograph (3) and Harper et al. (26) for more extensive discussions of this issue.) The health disparity indices (SII, RII, and the Health Concentration Index) can be used to compare health inequalities across time, geographic regions, and other subgroups to help monitor health disparities and public health efforts to reduce them.

A limitation of this work is that in the confidence interval estimation, groups were assumed to be mutually exclusive. Thus, categorizing a continuous measure (such as percentage of FRL participation in schools) into groups may involve some subjectivity, and misclassification bias is highest at the category cutpoints. In these analyses, we categorized percentage of FRL participation as ≥75%, 50–<75%, 25–<50%, and <25%, as in an Alameda County Public Health Department report (27). As a sensitivity analysis, we also categorized percentage of FRL participation based on quartiles of the distribution in the 186 schools, which yielded similar findings (not shown). The 1997 revision of US Office of Management and Budget Statistical Policy Directive 15 (28) stipulates that US federal data collection procedures allow respondents to indicate all racial designations that apply to them. Thus, to apply these methods for SII and RII, the mutually exclusive group of “more than 1 race” would have to be used instead of counting individuals as belonging to multiple groups. In future research, investigators could develop methods of allowing non-mutually exclusive multiple race selections by considering covariances among groups in the calculations, but a sufficient number of people with the various combinations would probably be needed. Furthermore, in additional work, researchers could incorporate covariate adjustments into calculation of the health disparity indices.

We provided both Taylor series linearization and a rescaled bootstrap approach for confidence interval estimation. We recommend using both methods, if possible, to assess potential asymmetry. The rescaled bootstrap method is more computationally intensive, but with this survey (n = 21,399 in 186 clusters), the program ran in only 39 minutes on a Dell Inspiron 600M laptop personal computer (1.6 GHz and 640 megabytes of random-access memory; Dell Computer Corporation, Round Rock, Texas) for 100 bootstrap samples, and it ran in 224 minutes for 500 bootstrap samples. The rescaled bootstrap method with 500 bootstrap samples produced narrower confidence intervals.

![Figure 5. Simulation illustrating the effect of disease prevalence (proportion) on the Health Concentration Index disparity measure with both rescaled bootstrap and Taylor series linearization (TSL) 95% confidence interval estimates for school percentage of participation in free/reduced-price lunch programs (≥75%, 50–<75%, 25–<50%, or <25%). Bars, 95% confidence interval.](image-url)
ACKNOWLEDGMENTS

Author affiliations: Center to Address Disparities in Children’s Oral Health, Division of Oral Epidemiology and Dental Public Health, School of Dentistry, University of California, San Francisco, San Francisco, California (Nancy F. Cheng, Stuart A. Gansky) and School of Public Health, University of California, Berkeley, Berkeley, California (Pamela Z. Han).

Funding was provided by the National Institute for Dental and Craniofacial Research (grant R03DE018116). The California Oral Health Needs Assessment of Children 2004–2005 (also known as the California Smile Survey) was conducted under the auspices of the Dental Health Foundation, with support from the Association of State and Territorial Dental Directors, the California Dental Association Foundation, First 5 California, the California Endowment, and the Health Services and Resources Administration (US Department of Health and Human Services).

The authors thank Dr. Kathy Phipps of Oregon Health and Science University and Wynne Grossman of the Dental Health Foundation for providing technical assistance regarding the design and conduct of the California Oral Health Needs Assessment of Children 2004–2005.

A SAS macro for using these health disparity indices and calculating confidence intervals is freely available from the University of California, San Francisco Center to Address Disparities in Children’s Oral Health (http://www.ucsf.edu/cando/resources.html#software).

Conflict of interest: none declared.

REFERENCES

APPENDIX

Rescaled Bootstrap Confidence Interval Method

The rescaled bootstrap method of Rao et al. (24) for estimating the confidence interval for an estimator \( \hat{\theta} \) of \( C \), the Health Concentration Index, is as follows.

1. Independently in each stratum \( h \), draw a simple random sample of \( m_h^{(b)} \) clusters with replacement from the \( n_h \) sample clusters. Let \( t_{hi}^{(b)} \) be the number of times that sample cluster \( hi \) (e.g., school \( i \) within geographic stratum \( h \)) is selected in the bootstrap sample \( b \), with \( b = 1, 2, \ldots, B \), where \( B \) is a large number such as 500. If sample cluster \( hi \) is not selected, then \( t_{hi}^{(b)} = 0 \).

2. Define the bootstrap weights for unit \( k \) in sample cluster \( hi \) by rescaling the survey weights \( w_{hik} \):

\[
\begin{align*}
    w_{hik}^{(b)} &= \left[ \left( 1 - \frac{m_h}{(n_h - 1)} \right)^{1/2} + \frac{m_h}{(n_h - 1)} \right]^{1/2} \left( \frac{n_h}{m_h} \right)^{t_{hi}^{(b)}} w_{hik}.
\end{align*}
\]

If \( m_h \leq n_h - 1 \), then the bootstrap weights \( w_{hik}^{(b)} \) are all positive. To simplify calculation of bootstrap weights, \( m_h = n_h - 1 \) is often used to reduce the above formula to

\[
\begin{align*}
    w_{hik}^{(b)} &= \left( \frac{n_h}{m_h} \right)^{t_{hi}^{(b)}} w_{hik}.
\end{align*}
\]

3. Calculate \( \hat{\theta}^{(b)} \), the bootstrap estimator of \( \theta \), by replacing the survey weights \( w_{hik} \) with the bootstrap weights \( w_{hik}^{(b)} \) in the formula for \( \hat{\theta} \).

4. Independently replicate steps 1–3 \( B \) times and calculate the corresponding estimates \( \hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)} \).

Assuming a normal distribution, the 95% bootstrap confidence interval is obtained from the 2.5th and 97.5th percentiles of the distribution of \( \hat{\theta}^{(b)} \).