**Appendix to “State-level Estimates of Childhood Obesity Prevalence in the United States Corrected for Report Bias” by Long et al.**

**Childhood Obesity Intervention Cost Effectiveness Study (CHOICES) Matching Process**

Statistical matching techniques allow researchers to combine data from distinct datasets that are sampled from the same underlying population, but do not contain any individual identifiers in common.[1](#_ENREF_1) For this project, we used statistical matching techniques to combine data on distinct individuals from the 2010 U.S. Census, the American Community Survey, the National Survey of Children’s Health (NSCH), and the National Health and Nutrition Examination Survey (NHANES).

The matching process described in this appendix was conducted using Java, a compiled programming language. The detail in the paper and appendix on the matching methods are intended to support replication by a skilled programmer in Java or other programming languages familiar with the theory and techniques discussed in textbooks on these methods.[1](#_ENREF_1) The methods can also be applied to related height and weight bias correction needs, as was done in an analysis of the healthcare costs associated with obesity and severe obesity using data from the Medical Expenditure Panel Survey by Wang et al.[2](#_ENREF_2)

We randomly sampled with replacement populations of 30 million individuals at the census tract level from the 2010 U.S. Census, which yielded 6.4 million individuals 2-17 years of age. Fifty populations were independently sampled in order to estimate variation in population estimates due to the random sampling process used in the matching algorithm.

To generate bias-corrected estimates of BMI at the state level using statistical matching, the random sample of 6.4 million children and adolescents from the 2010 U.S. Census was first matched by randomly sampling individuals in the 5-year ACS with replacement proportional to sampling weights within census tract, age, sex, and race/ethnicity subgroups to assign household income to each individual. The new dataset was then matched within state, age, sex, race/ethnicity and household income subgroups to the pooled 2003/2007 NSCH sample using random sampling with replacement of individuals in the NSCH sample proportional to sampling weights to assign parent-reported height and weight. Parent reported height and weight were converted to national-level percentiles within demographic subgroups in the sample-weighted NSCH dataset. Household income was reported as a categorical variable based on ranges of percentage poverty level. In order to be matched to the household income data from ACS, the NSCH percentage poverty level variable was converted to a household income range in dollars based on the year of the survey, state, and household size following the Department of Health & Human Services (HHS) guidelines. Crosswalks for linking race/ethnicity and household income across the Census/ACS, NSCH, and NHANES are included in Appendices E and F.

The resulting dataset combining inputs from the Census, ACS, and NSCH was then matched to individuals from the 2005-2010 NHANES on parent-reported height and weight percentiles from NSCH by sampling individuals with replacement proportional to sampling weights within age, sex, race/ethnicity, household income subgroups. For each individual in the NHANES sample, measured height and weight data are collected by NHANES trained technicians. Matching by height and weight was conducted in two ways. For adolescents aged 16-17 years, for whom self-reported height and weight were available from NHANES, individuals were matched within demographic strata based on sample-weighted self-reported height and weight percentiles within the NHANES dataset. Measured height and weight values from the matched individual in the donor dataset were then assigned to individuals in the synthesized dataset. Because children and adolescents 2-15 years of age did not self-report height and weight in NHANES, these individuals were matched within demographic strata based on their parent-reported height and weight percentiles from NSCH and their measured height and weight percentiles from NHANES. Measured height and weight were then assigned to individuals in the synthesized dataset.

Matching achieves greater precision through tightly-defined subgroups, such that all explainable variation in the imputed variables is captured in the process. However, the goal of precision must be balanced with the need to represent heterogeneity. A very tightly-defined subgroup may fail to adequately represent variation in the synthesized joint distribution and may result in no possible matches within a very narrow subgroup. Using a very broad subgroup definition may lead to inappropriate matches that fail to represent the explainable variation. To achieve this balance, we used dynamic subgroup definitions to achieve a minimum sample size, which we varied empirically to yield the desired balance between sample heterogeneity and matching precision.

We used age- and sex-specific thresholds that resulted in obesity prevalences statistically similar to NHANES. These thresholds were identified using a grid search. The target sample size for each age and sex group is included in Appendix G. If the subgroup sample was below the specified size, the matching restrictions were loosened until the specified sample size was met (see Appendix H). Within subgroup samples, percentile-matching bandwidths used to sample values of height and weight were initialized to zero and expanded in a similarly iterative way until a match was found.

Sampled measured height and weight were “smoothed” by randomly adding or subtracting up to 1% of the sampled value for each individual.[3](#_ENREF_3) The assigned smoothed height and weight values from matched individuals were used to derive BMI (kg/m2). BMI percentile was used to define overweight (≥85th and <95th percentile) and obesity status (≥95th percentile) based on the 2000 CDC growth charts.[4](#_ENREF_4)

**References**

1. D'Orazio M, Di Zio M, Scanu M. Statistical matching : theory and practice. Chichester, England ; Hoboken, NJ: Wiley; 2006. x, 256 p. p.

2. Wang YC, Pamplin J, Long MW, Ward ZJ, Gortmaker SL, Andreyeva T. Severe Obesity In Adults Cost State Medicaid Programs Nearly $8 Billion In 2013. Health Aff (Millwood) 2015;34(11):1923-31.

3. Silverman B, Young G. The Bootstrap: To Smooth or Not to Smooth? Biometrika 1987;74(3):469-79.

4. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC Growth Charts for the United States: methods and development. Vital and health statistics Series 11, Data from the national health survey 2002(246):1-190.

**Appendix A. Comparison of mean body mass index (BMI) by age and sex from parent-reported 2003/2007 National Survey of Children’s Health (NSCH) (n=133,213) and objectively measured 2005-2010 National Health and Nutrition Examination Survey (NHANES) (n=9,377)**

**Appendix B. Comparison of mean height, weight, and BMI by age and sex from 2003-2008 National Survey of Children’s Health (NSCH) (n=133,213) and 2005-2010 National Health and Nutrition Examination Survey (NHANES) (n=9,377)**

|  |
| --- |
| HEIGHT (cm) - Males |
| Age Group | **NHANES Mean (95% CI)** | **NSCH Mean (95% CI)** | **df** | **t** | **p** |
| 2-5 | 102.98 (102.35-103.62) | 99.09 (98.60-99.58) | 10,304 | -9.50 | <0.001 |
| 6-9 | 128.19 (127.53-128.84) | 124.26 (123.74-124.78) | 10,048 | -9.19 | <0.001 |
| 10-13 | 152.11 (151.29-152.94) | 150.70 (150.28-151.13) | 21,420 | -2.97 | 0.003 |
| 14-17 | 173.46 (172.86-174.05) | 174.21 (173.92-174.49) | 19,573 | 0.74 | 0.457 |
| HEIGHT (cm) - Females |
| Age Group | **NHANES Mean (95% CI)** | **NSCH Mean (95% CI)** | **df** | **t** | **P** |
| 2-5 | 101.50 (100.86-102.14) | 97.97 (97.47-98.48) | 9,765 | -8.46 | <0.001 |
| 6-9 | 128.35 (127.68-129.01) | 123.74 (123.12-124.35) | 9,702 | -9.95 | <0.001 |
| 10-13 | 152.68 (151.98-153.38) | 149.91 (149.46-150.35) | 20,483 | -6.52 | <0.001 |
| 14-17 | 162.34 (161.78-162.91) | 163.35 (163.13-163.57) | 24,197 | 3.24 | 0.001 |
| WEIGHT (kg) - Males |
| Age Group | **NHANES Mean (95% CI)** | **NSCH Mean (95% CI)** | **df** | **t** | **p** |
| 2-5 | 17.55 (17.30-17.81) | 17.96 (17.79-18.13) | 10,304 | 2.57 | 0.01 |
| 6-9 | 29.27 (28.67-29.88) | 29.21 (28.91-29.51) | 10,048 | -0.19 | 0.85 |
| 10-13 | 48.46 (47.35-49.57) | 47.76 (47.31-48.21) | 21,420 | -1.14 | 0.256 |
| 14-17 | 71.24 (69.80-72.68) | 69.59 (69.09-70.09) | 26,560 | -2.12 | 0.034 |
| WEIGHT (kg) - Females |
| Age Group | **NHANES Mean (95% CI)** | **NSCH Mean (95% CI)** | **df** | **t** | **p** |
| 2-5 | 16.91 (16.60-17.23) | 17.10 (16.96-17.24) | 9,765 | 1.06 | 0.288 |
| 6-9 | 29.38 (28.77-29.99) | 28.98 (28.63-29.34) | 9,702 | -1.11 | 0.266 |
| 10-13 | 49.75 (48.65-50.85) | 46.44 (46.03-46.85) | 20,483 | -5.54 | <0.001 |
| 14-17 | 62.19 (61.09-63.29) | 59.29 (58.94-59.63) | 24,197 | -4.95 | <0.001 |
| BMI (kg/m2) - Males |
| Age Group | **NHANES Mean (95% CI)** | **NSCH Mean (95% CI)** | **df** | **t** | **p** |
| 2-5 | 16.39 (16.28-16.51) | 18.95 (18.71-19.19) | 10,304 | 18.93 | <0.001 |
| 6-9 | 17.51 (17.28-17.73) | 19.19 (18.98-19.40) | 10,048 | 10.77 | <0.001 |
| 10-13 | 20.60 (20.26-20.94) | 20.89 (20.73-21.04) | 21,420 | 1.52 | 0.129 |
| 14-17 | 23.52 (23.11-23.93) | 22.83 (22.69-22.98) | 26,560 | -3.10 | 0.002 |
| BMI (kg/m2) - Females |
| Age Group | **NHANES Mean (95% CI)** | **NSCH Mean (95% CI)** | **df** | **t** | **p** |
| 2-5 | 16.24 (16.10-16.39) | 18.46 (18.27-18.65) | 9,765 | 18.2 | <0.001 |
| 6-9 | 17.52 (17.29-17.76) | 19.21 (18.95-19.47) | 9,702 | 9.47 | <0.001 |
| 10-13 | 21.05 (20.69-21.42) | 20.61 (20.44-20.77) | 20,483 | -2.18 | 0.03 |
| 14-17 | 23.56 (23.16-23.95) | 22.21 (22.08-22.34) | 24,197 | -6.38 | <0.001 |

**Appendix C. Comparison of BMI distributions by age and sex from NHANES 2005-2010 and CHOICES model**

**Appendix D. Childhood overweight and obesity prevalence by state in 2005-2010 estimated by CHOICES Model**



**Appendix E. Crosswalk for Linking Race/Ethnicity Codes across the Census/ACS, National Survey of Children’s Health (NSCH) and the National Health and Nutrition Examination Survey (NHANES)**

|  |  |  |
| --- | --- | --- |
| **Census/ACS** | **NSCH** | **NHANES** |
| White, Non-Hispanic | White, Non-Hispanic | White, Non-Hispanic |
| Black or African American, Non-Hispanic | Black or African American, Non-Hispanic | Black, Non-Hispanic |
| American Indian and Alaska Native, Non-Hispanic | Other, Non-Hispanic | Other Race, including Multi-Racial |
| Asian, Non-Hispanic |
| Native Hawaiian and Other Pacific Islander, Non-Hispanic |
| Other, Non-Hispanic |
| Two or more races, Non-Hispanic | Two or more races, Non-Hispanic |
| Hispanic | Hispanic | Mexican American |
| Other Hispanic |

**Appendix F. Conversion of Ratio of Household Income to Federal Poverty Level to Household Income from National Survey of Children’s Health (NSCH)**

|  |  |  |
| --- | --- | --- |
| Household Size | Poverty Threshold (2003) | Poverty Threshold (2007) |
| Continental US | Alaska | Hawaii | Continental US | Alaska | Hawaii |
| 1 | $8,980 | $11,210 | $10,330 | $10,210 | $12,770 | $11,750 |
| 2 | $12,120 | $15,140 | $13,940 | $13,690 | $17,120 | $15,750 |
| 3 | $15,260 | $19,070 | $17,550 | $17,170 | $21,470 | $19,750 |
| 4 | $18,400 | $23,000 | $21,160 | $20,650 | $25,820 | $23,750 |
| 5 | $21,540 | $26,930 | $24,770 | $24,130 | $30,170 | $27,750 |
| 6 | $24,680 | $30,860 | $28,380 | $27,610 | $34,520 | $31,750 |
| 7+ | $27,820 | $34,790 | $31,990 | $31,090 | $38,870 | $35,750 |

Poverty level categories (defined as a range of ratios of household income to poverty threshold) were multiplied by the poverty threshold in order to calculate ranges of household income.

**Appendix G. Optimum Subgroup Sample Size by Age and Sex**

|  |  |  |
| --- | --- | --- |
| Age Group | Males | Females |
| 2-5 | 125 | 125 |
| 6-9 | 125 | 75 |
| 10-13 | 200 | 200 |
| 14-17 | 200 | 125 |

**Appendix H. Dynamic Subgroup Matching Definitions to Achieve Optimum Subgroup Sample Size**

**Census/ACS to NSCH**

|  |  |  |
| --- | --- | --- |
| **Iteration** | **Matching Group** | **Cumulative % Matched** |
| 0 | (Exact Match) | 60.62% |
| 1 | Income +/- 5,000 | 76.69% |
| 2 | Income +/-10,000 | 79.88% |
| 3 | Adjacent age | 89.53% |
| 4 | Income +/- 15,000 | 91.27% |
| 5 | Income +/- 20,000 | 92.17% |
| 6 | Income +/- 25,000 | 92.89% |
| 7 | Income +/- 35,000 | 93.86% |
| 8 | Income +/- 45,000 | 94.61% |
| 9 | Income +/- 55,000 | 95.05% |
| 10 | All Income | 96.22% |
| 11 | All Ethnicities | 99.82% |
| 12 | All Races | 100% |

**NSCH to NHANES**

|  |  |  |
| --- | --- | --- |
| **Iteration** | **Matching Group** | **Cumulative % Matched** |
| 0 | (Exact Match) | 38.00% |
| 1 | Income Range 1 | 57.77% |
| 2 | Income Range 2 | 60.47% |
| 3 | Income Range 3 | 63.92% |
| 4 | Income Range 4 | 66.72% |
| 5 | Income Range 5 | 69.77% |
| 6 | Income Range 6 | 71.93% |
| 7 | Income Range 7 | 73.63% |
| 8 | All Income | 88.84% |
| 9 | All Ethnicities | 100% |