2019 Add Health Wave I Documentation

Wave I County Health, Mobility, and Tobacco Tax Data Documentation

Report prepared by

Daniel W. Belsky

Lauren Gaydosh

Max Reason

Taylor W. Hargrove

UNC CAROLINA POPULATION CENTER

CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

Add Health is supported by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations.

https://doi.org/10.17615/thjx-1297

Suggested Citation

Belsky, Daniel W., Lauren Gaydosh, Max Reason, and Taylor W. Hargrove, 2019. The National Longitudinal Study of Adolescent to Adult Health (Add Health), Wave I County Health, Mobility, and Tobacco Tax Data Documentation. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. https://doi.org/10.17615/thjx-1297

Acknowledgments

The creation of this database was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development R21 HD095448, National Institute of Health: P2C HD050924, and the Jacobs Foundation Research Fellowship Program 2015-1170-09. We are grateful to Tim Monbureau, Joyce Tabor, Kathleen Mullan Harris, Robert Hummer, Sarah Dean, and Maria Marrufo, in addition to staff from the Add Health project and the Carolina Population Center, in helping us produce this data base.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Introduction

The Wave I County Health and Mobility database summarizes the socioeconomic, health, and mobility characteristics of the environments in which Add Health participants were living at the time of their Wave I interview. County-level data describe (1) levels of and trends in chronic disease (hypertension, type-2 diabetes) and health risk behaviors (obesity, smoking, alcohol use); and (2) economic opportunity and inequality. This contextual database permits innovative research that investigates how place influences health, behavior, and social outcomes during the transition from adolescence to adulthood, thereby, enhancing studies of the determinants and sequelae of socio-geographic mobility. Comprehensively, the database provides indicators of contextual conduciveness to the socio-geographic and health-geographic mobility of Add Health participants from Wave I to Waves IV and V.

Data

The following is a list of data that were collected from secondary data sources and merged to Wave I of Add Health. These variables are available at the county or state level. Data were matched to the county or state that the Add Health respondent was living in at the time of the Wave I interview. Data were matched to respondents so as to ensure that these contextual variables correspond as closely as possible to the year in which the Add Health respondents were interviewed at Wave I (1994/1995).

Life Expectancy and Mortality Risk

Data on life expectancy and mortality risk come from the Institute for Health Metrics and Evaluation (IHME). Annual county-level life tables were constructed using small area estimation methods from de-identified death records from the National Center for Health Statistics (NCHS), and population counts from the US Census Bureau, NCHS, and the Human Mortality Database.

County level life expectancy (1995) C1LE001
County level mortality risk age 0-5 (1995) C1LE002
County level mortality risk age 5-25 (1995) C1LE003
County level mortality risk age 25-45 (1995) C1LE004
County level mortality risk age 45-65 (1995) C1LE005
County level mortality risk age 65-85 (1995) C1LE006

Citation: Lindgren et al. 2017 *JAMA Internal Medicine* Inequalities in Life Expectancy Among US Counties, 1980-2014 [1]

Diabetes Prevalence

Data on diabetes prevalence came from the IHME. IHME researchers used a two-stage modeling procedure. In the first stage, self-reported and biomarker data from National Health and Nutrition Examination Survey (NHANES) were used to build models for predicting true diabetes status, which were applied to impute true diabetes status for respondents in the Behavioral Risk Factor Surveillance System (BRFSS). In the second stage, small area models were fit to imputed BRFSS data to derive county-level estimates of diagnosed, undiagnosed, and total diabetes prevalence, as well as rates of diabetes diagnosis and effective treatment.

County level age-standardized diagnosed diabetes prevalence: Total (1999) C1DI001T
County level age-standardized diagnosed diabetes prevalence: Male (1999) C1DI001M
County level age-standardized diagnosed diabetes prevalence: Female (1999) C1DI001F
County level age-standardized undiagnosed diabetes prevalence: Total (1999).....C1DI002T
County level age-standardized undiagnosed diabetes prevalence: Male (1999)..... C1DI002M
County level age-standardized undiagnosed diabetes prevalence: Female (1999) C1DI002F
County level age-standardized total diabetes prevalence: Total (1999) C1DI003T
County level age-standardized total diabetes prevalence: Male (1999) C1DI003M
County level age-standardized total diabetes prevalence: Female (1999) C1DI003F
County level age-standardized diabetes awareness: Total (1999) C1DI004T
County level age-standardized diabetes awareness: Male (1999) C1DI004M
County level age-standardized diabetes awareness: Female (1999) C1DI004F
County level age-standardized diabetes control: Total (1999).....C1DI005T
County level age-standardized diabetes control: Male (1999)..... C1DI005M
County level age-standardized diabetes control: Female (1999) C1DI005F

Citation: Dwyer-Lindgren et al. 2016 *Diabetes Care* Diagnosed & Undiagnosed Diabetes Prevalence by County in the US 1999-2012 [2]

Drinking Patterns

Data on drinking patterns came from the IHME. IHME researchers applied small area models to BRFSS data on self-reported drinking, incorporating spatial and temporal smoothing.

County level age-standardized prevalence of any drinking: Total (2002) C1DR001T
County level age-standardized prevalence of any drinking: Male (2002) C1DR001M
County level age-standardized prevalence of any drinking: Female (2002)..... C1DR001F
County level age-standardized prevalence of binge drinking: Total (2002) C1DR003T
County level age-standardized prevalence of binge drinking: Male (2002) C1DR003M
County level age-standardized prevalence of binge drinking: Female (2002) C1DR003F

Citation: Dwyer-Lindgren et al. 2015 *AJPH* Drinking Patterns in US Counties from 2002-2012 [3]

Smoking Patterns

Data on smoking patterns came from the IHME. IHME researchers used data on 4.7 million adults age 18 and older from the BRFSS from 1996 to 2012. They derived cigarette smoking status from self-reported data in the BRFSS and applied validated small area estimation methods to generate estimates of current total cigarette smoking prevalence and current daily cigarette smoking prevalence.

County level prevalence of people who currently smoke: Total (1996)..... C1SM001T
County level prevalence of people who currently smoke: Male (1996)..... C1SM001M
County level prevalence of people who currently smoke: Female (1996) C1SM001F
County level prevalence of people who currently smoke daily: Total (1996) C1SM002T
County level prevalence of people who currently smoke daily: Male (1996) C1SM002M
County level prevalence of people who currently smoke daily: Female (1996)..... C1SM002F

Citation: Dwyer-Lindgren et al. 2014 *Population Health Metrics* Cigarette Smoking Prevalence in US Counties 1996-2012 [4]

Physical Activity and Obesity

Data on physical activity and obesity came from the IHME. Body mass index (BMI) is calculated from self-reported weight and height in BRFSS, adjusting for self-reporting bias using NHANES. Physical activity—both any physical activity and physical activity meeting recommended levels—is calculated from self-reported data in BRFSS. To generate estimates of obesity and physical activity prevalence for each county, IHME researchers used validated small area estimation methods.

County level prevalence of obesity: Male (2001) C1OP001M
County level prevalence of obesity: Female (2001)..... C1OP001F
County level prevalence of physical activity: Male (2001) C1OP002M
County level prevalence of physical activity: Female (2001)..... C1OP002F

Citation: Dwyer-Lindgren et al. 2013 *Population Health Metrics* Prevalence of physical activity and obesity in US counties, 2001–2011: a road map for action [5]

Measures of Mobility

Data on measures of mobility come from the Equality of Opportunity Project. These data characterize US counties in terms of resident socioeconomic mobility. The Equality of Opportunity Project draws on federal income tax data to create measures of mobility.

County relative mobility – slope from OLS regression of child rank on parent rank within each county in core sample using baseline income definitions. Correlation of the percentile rank in the national income distribution for children (based on average incomes between 2010 and 2012 for the 1980-1982 birth cohort) and their parents (whose income was measured over 1996-2000).

County absolute mobility – expected rank of children whose parents are at the 25th percentile of the national income distribution based on the rank-rank regression.

Causal effect of county of childhood residence on adult household income –measured as the percentage gain or loss in income at age 26 caused by spending one additional year of childhood in a given county relative to the national mean for children born to a family earning an income of approximately $30,000 (the 25th percentile of the income distribution). Children were assigned to a county based on their location at age 16 (no matter where they live as adults), so that their location represents where they grew up.

County Gini coefficient – a measure of the amount of parental income inequality within commuting zones in the US. The following equation is used to compute the Gini coefficient:

$$Gini = \frac{2}{\overline{X_c}} Cov(X_{ic}, P_{ic}),$$

where $\overline{X_c}$ is the mean family income (for 1996-2000) of parents in CZ *c*, and $Cov(X_{ic}, P_{ic})$ is the covariance between income level (X_{ic}) and percentile rank (P_{ic}) of parents in CZ *c*.

County relative mobility	C1EC001
County absolute mobility	C1EC002
Causal effect of county of childhood residence on adult household income at p25: Total	C1EC003T
Causal effect of county of childhood residence on adult household income at p25: Male	C1EC003M
Causal effect of county of childhood residence on adult household income at p25: Female	C1EC003F
County Gini coefficient	C1EC004

Citations: Chetty & Hendren 2018 The Effects of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects [6]

Chetty & Hendren 2018 The Effects of Neighborhoods on Intergenerational Mobility II: County-level Estimates [7]

Chetty, Hendren, Kline & Saez 2014 Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States [8]

Tax Burden on Tobacco

Data come from the Tax Burden on Tobacco dataset.

State level cigarette tax per pack, in dollars (1994) S1CT001

Citation: Orzechowski and Walker, 2016 Tax Burden on Tobacco [9]

Missing Codes

The final digit of the missing codes indicates the reason for which they are missing. Missing codes that end in 2 (Ex. 92, 992) denote that information for that variable was not available in the source dataset. Missing codes that end in 8 (Ex. 98, 998) denote respondents in Add Health who lack the geocodes necessary for merging respondent locations to the various source data.