

Small Area Estimation



State and local agencies often face difficulties when designing health promotion and prevention activities because they lack local, area-specific data pertaining to counties, groups of counties, health service areas, and other administrative units. Most states and local areas do not possess the economic resources or infrastructure to collect this volume of data via a direct survey approach. RTI International leverages advances in statistics and computing power to offer a viable, affordable alternative that permits the production of valid and reliable estimates for small areas.

Overview

Small area estimation (SAE) is the process of using statistical models to link national or state survey outcome variables, such as disease indicators, to local area predictors, such as county demographic and socioeconomic variables, so that prevalence rates for small areas can be predicted. Social indicator variables such as age, race/ethnicity, gender, education, income, family structure, and employment status are commonly used to define high-risk subpopulations for targeting health promotion and disease prevention. Relating health status, behavior, and disease prevalence statistics for small areas like counties to these demographic and socioeconomic predictors provides a direct calibration of the indicators to the outcomes of interest. SAE methods can be applied to cases where the number of area-specific sample observations is not large enough to produce reliable direct estimates. The areas may be defined by geographic domains such as state or county or by sociodemographic characteristics such as income, race, age, or gender subgroups.

Project Highlights

RTI has developed an innovative survey-weighted hierarchical Bayes (SWHB) solution for fitting mixed logistic models that has been implemented successfully on data from the Substance Abuse and Mental Health Services Administration (SAMHSA) National Household Survey on Drug Use and Health (NSDUH). The annual NSDUH has a 50-state design with an independent, multistage area probability sample for each of the 50 states and the District of Columbia. State-level small area estimates have been produced annually since 1999 by pooling 2 years of NSDUH data for over 20 binary outcomes related to substance use (e.g., marijuana, cocaine), treatment (needing but not receiving treatment for alcohol or drug problems), and mental health (depression).

RTI has also used NSDUH data to produce sub-state estimates (typically for groups of counties or Census tracts) for SAMHSA by pooling data from three consecutive survey years. These sub-state areas were defined by state officials and based on treatment planning areas specified by the states in their applications for the Substance Abuse

Small Area Estimation

Prevention and Treatment Block Grant administered by SAMHSA. The goal of the sub-state SAEs was to display the geographic variation of substance use prevalence for areas that states would find useful for treatment planning purposes.

RTI's SWHB solution has several benefits over other hierarchical Bayes solutions obtained from commercially available software packages such as WinBUGS and MLwiN:

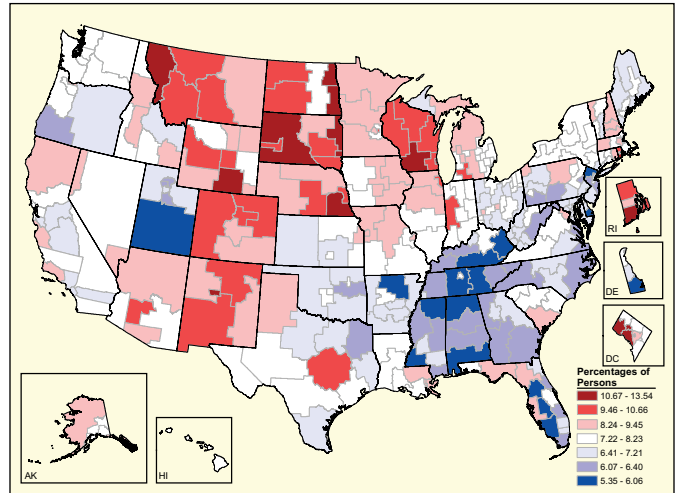
- SAEs for large sample areas are close to their design-based analogs; hence, they are robust against model misspecification.
- National aggregates (e.g., state-level SAEs) of the SAEs are design-consistent and approximately self-calibrated to the robust design-based national estimates.
- The SWHB methodology allows for the use of person- or unit-level predictors as opposed to aggregate-level predictors only. As a result, SAEs based on the SWHB method are internally consistent and also more precise than solutions obtained from aggregate-level small area models.

RTI statisticians have applied SAE techniques to other projects as well:

- RTI worked on a research grant awarded to the UNC Center for Health Statistics Research by the Centers for Disease Control and Prevention to produce county-level prevalence estimates of diabetes and smoking for all counties in North Carolina using Behavioral Risk Factor Surveillance System (BRFSS) data. These estimates were used to identify high-risk areas (counties or groups of counties) at which prevention and other health intervention programs can be directed. For this purpose, data from seven consecutive BRFSS surveys (1996 to 2002) were used.
- SAE methodology was used to produce state-level prevalence rates for high daily person miles of travel and associated prediction intervals for all 50 states and the District of Columbia using the 2001 National Household Transportation Survey. A high daily person miles of travel was defined as “more than 87.5 miles traveled in a day” (the 90th percentile for daily person miles traveled). This project was funded by the Bureau of Transportation Statistics.

GIS Mapping

Mapping is the cornerstone for communicating geographic information. RTI uses geographic information system (GIS) software to create graphical applications such as mapping disease incidence or substance use rates. The combined power of statistical expertise and geographic science provides outstanding value to our clients.



This map produced by the GIS team for SAMHSA presents sub-state level SAEs for dependence on or abuse of alcohol among persons aged 12 or older based on 2002-2004 combined NSDUHs.

More Information

Akhil Vaish
919.541.6719
avaish@rti.org

RTI International
3040 E Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194 USA

RTI 7700 R1 0715



RTI International is one of the world's leading research institutes, dedicated to improving the human condition by turning knowledge into practice. Our staff of more than 3,700 provides research and technical services to governments and businesses in more than 75 countries in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, international development, economic and social policy, energy and the environment, and laboratory testing and chemical analysis. For more information, visit www.rti.org.

RTI International is a registered trademark and a trade name of Research Triangle Institute.