

Innovate.
Evolve.
Lead.

Survey research is ever-changing and as experts and leaders in the field, we have continued to push the boundaries to get to the best insights for our clients.

Explore how we are still changing the game through survey innovation.



1970s

1980s

1990s

2000s

2010s+

Still Changing the Game Through Survey Innovation - Timeline

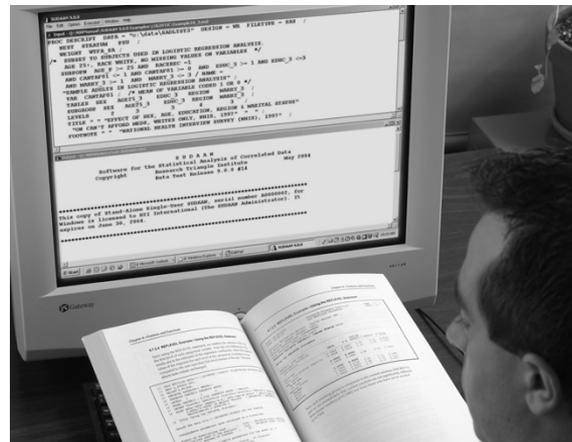
▶ 1972 | **SUDAAN®**

SUDAAN is an internationally recognized statistical software package that provides efficient and accurate analysis of data from complex studies. It is the only software offering all three popular robust variance estimation methods in its descriptive and modeling procedures:

- ▶ Taylor series linearization (generalized estimating equation for regression models)
- ▶ Jackknife (with or without user-specific replicate weights)
- ▶ Balance repeated replication (BRR)

▶ 1974 | **Taylor Series Variance Estimators for Complex Survey Estimates | Folsom**

RTI pioneered the correct calculation of sampling variances while considering design complexities like stratification and clustering. Appropriate software (i.e., SUDAAN) was developed and has been updated and improved on a regular basis. The impacts puts more easily accessible methods for incorporating design complexities in the analysis of social science data into the hands of researchers.



1970s

▶ **1976 | Double Sampling Difference Estimators | Lessler**

Double sampling is used to collect additional information to derive a more accurate difference estimator. The statistical methodology underlying this method has demonstrated to users how to derive difference estimators under a double sampling scenario.

1980s

1990s

▶ **1979 | Probability Proportional to Size (PPS) Sampling | Chromy**

RTI's Dr. James R. Chromy developed the theory and computational algorithm for selecting samples sequentially with PPS and with minimum replacement, and he developed a computer algorithm for efficient sample allocation that minimizes total survey cost subject to satisfying multiple variance constraints.

2000s

2010s+

1970s

1980s

1990s

2000s

2010s+

- ▶ **1980 | Improved Weighted Sequential Hot Deck Algorithm | Cox**

The Weighted Sequential Hot Deck Imputation (WSHD) differs from the standard hot deck in that the sample weights are used in selection of a donor. WSHD gives the resulting weighted distribution of imputed variables closer to the distribution of original variables of item responders, which is a main measure of the quality of imputation. In addition, the number of times the donor is selected for imputation is proportional to its weight. Macros to conduct WSHD imputation were developed by Iannacchione (1982). RTI has recently included the WSHD within a macro developed to implement cyclical tree-based imputation.



1970s

1980s

1990s

2000s

2010s+

- ▶ **1991 | Calibration Weighting with the General Exponential Model**

Since 1991, RTI has been at the forefront of developing weighing techniques for handling unit nonresponse in complex sample surveys and measuring the impact on variance resulting from those and other calibration weighting methods. By using a variant of the general exponential model available in the SUDAAN 11[®] software, a user can treat unit nonresponse as a logistic or truncated logistic model of covariates in a manner that produces estimates with less variance than using analogous maximum-likelihood methods. With calibration weighting and SUDAAN, nonresponse can be either missing at random or not missing at random (i.e., a function of variables, including the survey variable).



1970s

1980s

1990s

2000s

2010s+

▶ **1994 | Small Area Estimation (SAE) | National Survey on Drug Use and Health (NSDUH)**

SAE, the process of using statistical models to link national or state survey outcome variables to local area predictors to predict small area prevalence rates, was implemented in our work with NSDUH in 1994. Using SAE, we were able to display geographic variation of substance use prevalence for areas that states could use for more precise treatment planning purposes.

▶ **1995 | Audio Computer Assisted Self-Interviewing (ACASI) | National Survey of Adolescent Males**

The creation of ACASI allowed survey participants to now answer surveys via computer with headphones, eliminating data entry time and reducing room for error. ACASI improved the accuracy of data collected for surveys on sensitive topics, including abortion, drug use, and sexual behaviors and practices when compared with traditional survey administration.



1970s

1980s

1990s

2000s

2010s+

- ▶ **1996 | Latent Class Analysis (LCA) as a Survey Estimation Tool**

RTI's Paul Biemer's wrote *Latent Class Analysis of Survey Error*, which was the first book of its kind and explored 50+ different ways for using LCA and multimethod LCA in survey data analysis and error evaluation.

- ▶ **1999 | Predictive Mean Neighborhood (PMN) Imputation | Folsom, Grau, and Singh**

PMN imputation is a combination of two commonly used imputation methods: non-model-based nearest neighbor hot deck (NNHD) (Little & Rubin, 1987, p. 65) and a modification of Rubin's model-assisted predictive mean matching (PMM) method (Rubin, 1986), which matches a missing value to the observed value with the closest predicted mean. The PMN method enhances Rubin's PMM method, in that PMN can be applied to both discrete and continuous variables, either individually or jointly. PMN also enhances the NNHD method for discrete variables so that the distance function used to find neighbors is no longer in terms of the original predictor variables and therefore does not require arbitrary scaling. PMN also allows the user to guarantee the internal consistency of the post-imputation record by placing restrictions on the donor pool.

1970s

1980s

1990s

2000s

2010s+

- ▶ **1999 | Computer Audio Recorded Interviewing (CARI) | National Survey of Child and Adolescent Well-Being (NSCAW)**

CARI creates a digital record of the exchanges between interviewer and respondent from either telephone or in-person interviews. It provides a valuable opportunity to improve the quality of data collection, interviewer performance, and questionnaire design. Compared with other quality assurance methods, CARI saves time and money by reducing or eliminating the need for live monitoring, re-interviews, and labor-intensive verification.



- ▶ **1999 | Implemented a 50-State Sample Design to Enhance a National Design | NSDUH**

Annually, 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 (e.g., needing but not receiving treatment for alcohol or drug problems), and mental health (e.g., depression).

1970s

1980s

1990s

2000s

2010s+

- ▶ **2000 | Address-Based Sampling (ABS) | [Dallas Heart Study](#)**

ABS has emerged as a cost-effective sampling frame that offers high coverage of the U.S. household population for in-person, mail, telephone, and multimode surveys. ABS frames rely on residential addresses in the U.S. Postal Service (USPS) Computerized Delivery Sequence File provided by third-party vendors. In 2000, RTI conducted the first published evaluation of mailing addresses for use with an in-person household survey in the United States. The evaluation found that a sampling frame of mailing addresses provided nearly complete coverage of the Dallas, Texas, metropolitan area at approximately one-tenth the cost of traditional area sampling.

- ▶ **2004-2007 | Measures to Protect the Identity of Respondents from Inadvertent Exposures | MASSC**

RTI developed an innovative statistical disclosure control methodology called MASSC for creating public use micro data files. MASSC uses four steps: (1) **Micro Agglomeration**, (2) optimal probabilistic **Substitution**, (3) optimal probabilistic **Subsampling**, and (4) optimal sampling weight **Calibration**. MASSC simultaneously controls and optimizes disclosure risk and information loss in NSDUH. It has been used for data de-identification since the release of the 1999 NSDUH public use file. MASSC is undergoing some improvements in its application for higher data quality while minimizing disclosure risk.

1970s

1980s

1990s

2000s

2010s+



▸ **2006 | MicroPEM™ for Children's Gulf Coast Exposure | Children's Health After The Storm**

To better understand the potential environmental health risks associated with children living in temporary housing units or trailers following Hurricanes Katrina and Rita, RTI deployed the next generation of the MicroPEM, personal exposure monitoring device, to collect ambient data. Since it was first used in 2011, MicroPEM has helped researchers fill the gaps in our understanding of acute and chronic aerosol exposures, which were previously expensive and complex to monitor directly.

▸ **2007 | Document Manager (DocMan) | NSCAW**

DocMan is an RTI-developed system allowing paperless data collection that also fulfills data security needs

1970s

1980s

1990s

2000s

2010s+

- ▶ **2014 | Redesigned Sample to Improve Precision of Estimates for Older Adults and Reduce Unequal Weighting | NSDUH**

In advance of a planned National Survey on Drug Use and Health (NSDUH) sample redesign, several methods studies were conducted to optimize the design with respect to variance and cost while meeting other sample requirements (e.g., ensuring sufficient sample in each state and the District of Columbia to allow for state and substate estimation). As a result, the redesigned 2014 NSDUH allocated more sample to the older age groups to more accurately estimate drug use and related mental health measures among the aging population. In addition, the allocation of sample to states was closer to a proportional allocation to increase the efficiency of estimates (i.e., reduce unequal weighting).



1970s

1980s

1990s

2000s

2010s+

- ▶ **2015 | Modified Predictive Mean Neighborhood (modPMN) Imputation | Kott, Folsom**

The ModPMN imputation method was developed to improve upon the PMN method. Under modPMN, the hot-deck step for categorical variables is replaced with a stochastic process to “center” the imputed value as described in Singh, Grau, and Folsom (2004). Closely related variables are imputed in a cyclical approach to reduce the importance of imputation order and to simplify predictive mean vectors and constraints. Under modPMN, domain members are segregated into five bins based on their estimated propensity to respond and the prediction modeling and imputation step are performed separately within each bin. This promotes “double protection” as described in Kott and Folsom.

1970s

1980s

1990s

2000s

2010s+

- ▶ **2018 | Methods for Handling Missing Item Values in Regression Models Using the National Survey on Drug Use and Health**

This project involved creating a monograph that guides analysts interested in fitting regression models using complex survey data. The monograph includes a theoretical review of existing missing-data methods, an innovative simulation study that evaluates several of the more promising methods using NSDUH datasets, and a final chapter where the results of both the theoretical review and the simulation study are synthesized into guidance for analysts via decision trees.

- ▶ **2018 | Big Data Meets Survey Research Conference (BigSurv)**

RTI co-created the BigSurv conference, first held in October 2018 in Barcelona, Spain. During the 3-day event, attendees, panelists, and presenters engaged in a variety of discussions on the practical applications for employing big data and data science to improve the quality of statistics production. This was a first-of-its-kind event that will continue to be held every 2 years to connect survey researchers around the globe.



Visit our website for more information
on our work in survey research.

<https://www.rti.org/service-capability/surveys-and-data-collection>

