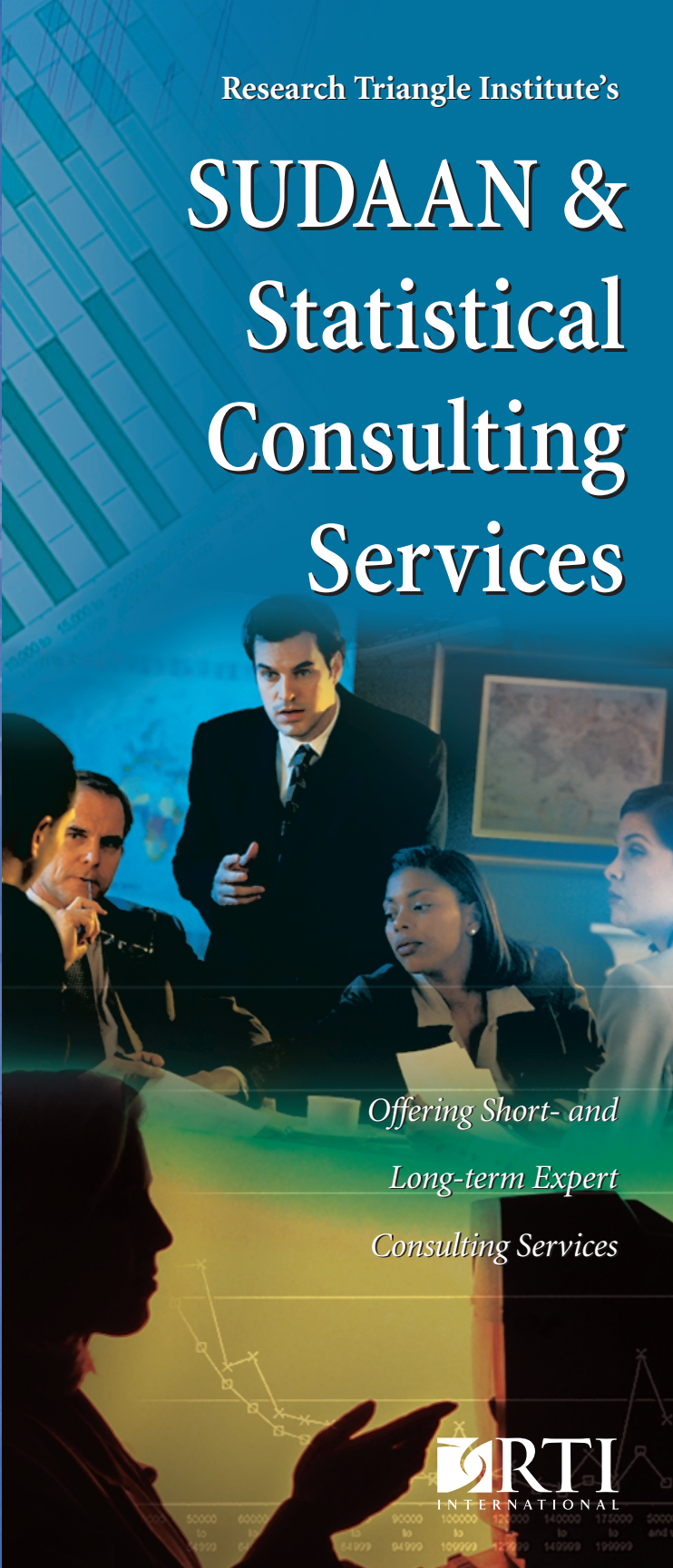
Research Triangle Institute's

SUDAAN & Statistical Consulting Services



Offering Short- and Long-term Expert Consulting Services

RTI INTERNATIONAL

RTI is committed to providing superior quality research and technical consulting services. We strive to make our clients comfortable with the statistical integrity and defensibility of our results and advice.

Contents

SUDAAN and Statistical Consulting Services

Offering Statistical Solutions for

- SUDAAN Programming and Results Interpretation
- Survey Design and Analysis
- Item Imputation Techniques
- Sample Weighting
- Small Area Estimation
- Environmental Statistics
- Health Outcomes Research
- Genetics and Genomics Research
- Bioinformatics
- Measuring Sampling & Nonsampling Errors
- Linear & Nonlinear Modeling, Diagnostics and Interpretation
- Analysis of Safety/Efficacy Clinical Trial Data
- Efficient and Timely Procedures for Collating Results for Electronic NDA submissions
- Efficient General Table Production Programming (MS Word, Word Perfect, HTML, PDF)

Research Triangle Institute (RTI) is proud to offer consulting services from our expert statisticians in SUDAAN programming and results interpretation, biostatistics, survey design and analysis, environmental statistics, pharmaceutical and health outcomes research, bioinformatics, sampling and nonsampling error measurement, and statistics for the behavioral and social sciences. Our staff of over 100 statisticians is committed to providing our clients with sound statistical designs and cost-effective approaches for answering research and statistical programming questions.

RTI recently developed this program to provide timely, responsive consulting services to our clients. Our consulting commitment can be as short as a few hours as long as several months, depending on the complexity of the request. In every case, RTI is committed to providing quality services. This commitment to quality, coupled with the diverse experience and knowledge of our staff, enables us to provide responses to a wide variety of issues that are statistically sound, defensible and computationally efficient.

Additional information on the SUDAAN and Statistics Consulting Program can be obtained by visiting http://www.rti.org/sudaan or by contacting the

SUDAAN Statistical Software Center
(919) 541-6602 E-Mail: sudaan@rti.org

Gertrude M. Cox Statistics Building, RTI Campus

Our Quality-Review Approach

RTI's Commitment to Quality

RTI's commitment to providing our clients superior quality research and technical services permeates all our work. We continue this tradition by ensuring that responses to even the smallest statistical consulting assignments undergo a formal quality review prior to the delivery of any product or service. Our goal is to not just respond to your analytic needs but also to help provide you with a thorough understanding of our response to your questions. Most importantly, we want you to feel comfortable with the statistical integrity and defensibility of our results and advice.

Five-step Quality-Review Process

Our quality-review approach consists of a basic five-step process:

1. A senior statistician reviews the request for statistical consultation, determines the optimal mix of staff, assembles a team, and assigns a primary statistician.
2. The primary statistician contacts the client to clarify the scope of work, estimate the level of work, and carefully review this with the client. We begin work once we have client approval.
3. A second, independent statistical reviewer reviews and formally approves the scope of work, analysis plan, and/or deliverables.
4. After the analysis is complete and any resulting tables, reports, programming code, etc. are formally approved by the statistical reviewer, we send the deliverables to the client.
5. Both the primary statistician and statistical reviewer follow up with the client to answer any questions.

Maintaining Confidentiality

Consistent with RTI's formal pledge of confidentiality, any analysis, datafile, and/or microdata in any form provided to us by a client or generated by the assigned consulting team, are considered strictly confidential. Data and results are never released nor reviewed by any individual outside the consulting team. Additionally, all data and results are destroyed or returned to the client at the end of the assignment.

SUDAAN Consultation

SUDAAN Software

SUDAAN® is a software product designed and developed by RTI statisticians for analyzing clustered data arising in many applications, including complex sample surveys, randomized experiments, and epidemiological studies. SUDAAN® offers procedures for descriptive statistics and for statistical modeling. The modeling procedures include both linear and multinomial-logistic models, and a procedure for fitting proportional hazards (Cox) regression models for time-to-event data. Additionally, SUDAAN is the only software package to offer all three popular robust variance estimation methods in its descriptive and modeling procedures:

- Taylor series linearization (GEE for regression models);
- Jackknife (with or without user-specified replicate weights) and
- Balanced repeated replication (BRR).

Together, these procedures and techniques provide a powerful variety of analytic tools not available in other software packages.

SUDAAN Consultation

RTI has a dedicated team of statisticians ready to respond to any SUDAAN consultation issue. Our SUDAAN consultation services can include

- Determining the appropriate SUDAAN variance estimation, design, and procedure options for a particular analysis.
- Writing the appropriate SUDAAN code for an analysis under consideration.
- Conducting the appropriate analyses using SUDAAN in conjunction with other software packages (e.g., SAS).
- Using SAS and SUDAAN to format specific SUDAAN results in client-specified tables (e.g., WordPerfect, MS Word, HTML, PDF) for inclusion in reports, press releases, manuscripts, web sites, etc.

Our clients can benefit from the efficiency and analytic appropriateness of SUDAAN without necessarily purchasing the software product. This is particularly beneficial for clients interested in using SUDAAN and the underlying methodology for unique, one-time, or limited duration projects.

Survey Statistics

RTI's unmatched capabilities in survey statistics enable us to consult with clients on virtually every aspect of survey design and analysis.

Sample Design. We have experts in determining appropriate sample design and selection techniques including stratification, clustering, multistage and multiphase sampling, PPS sampling, controlled selection, and multiple-frame sampling. Our experience with sample selection methods has included list sampling (e.g., businesses and Medicare or Medicaid recipients), telephone designs (RDD), area survey designs, and most recently, web-based sample designs.

Design Optimization. We pioneered many techniques that are now widely used, including the use of composite size measures to efficiently locate and interview members of rare populations and probability minimum replacement sampling. In addition, we have developed a methodology and software for determining the optimal sample size and sample allocation to achieve multiple precision constraints at minimum costs.

Questionnaire Design. Our integrated focus led us to develop innovative ways to study and reduce measurement errors in research projects. We are leaders in the development of cognitive laboratory methods for studying the survey response process, and we have conducted field experiments examining the effect of alternative measurement methods and data collection protocols on final estimates.

Experimental Design. We embed experimental designs into sample surveys to estimate differences in response rates and differences in measurement errors between alternative survey instruments and/or alternative modes of interviewing. We also develop experimental designs to characterize the performance of analytical laboratory procedures.

Pre-Analysis Techniques. We develop and utilize the most current and efficient pre-analysis statistical tools, including

- Use of model-based techniques to adjust for missing data, as well as more classical methods of imputation, including weighted and unweighted hot deck ("carry-over" procedure), mean imputation, and model-based imputation techniques.
- Multivariate item imputation techniques to account for the underlying correlation between an individual's responses to several questions.
- Model-based sample weight adjustments for efficient and quick nonresponse and post-stratification adjustment.

Analysis Techniques. We have extensive experience in various analysis techniques, including

- Determining the appropriate variance estimation and point estimation techniques.
- Developing appropriate contrast estimation methodologies to determine significance of differences between subgroups.
- Developing appropriate variance estimation techniques for multi-phase sample designs and multiple imputation.

Efficient Table Production Methodologies. We are particularly experienced at utilizing creative SAS/SUDAAN programming coupled with MS Word/WordPerfect Visual Basic macro programming to quickly and efficiently produce large numbers of publishable tables and graphs.

Research in Nonsampling Errors

Survey, clinical trial, and experimental data may contain errors as a result of data-collection and data-processing mistakes, nonresponse and missing data values, the biasing influences of interviewers and clinicians, respondent memory problems, and so on. Existing literature indicates that such nonsampling errors are often more important than sampling errors in the analysis and interpretation of results. Recognizing this, RTI has for many years maintained a productive and innovative staff of researchers with special expertise in the area of nonsampling error.

Compensating for Nonresponse. RTI researchers have made substantial efforts to increase response rates through better survey and experimental design, improved interviewer training and supervision, and the use of incentives and more private data collection modes. To deal with the problem of residual nonresponse, RTI's statisticians have developed new weight adjustment methods based on generalized exponential and logistic raking procedures. In the area of data imputation, RTI researchers have developed weighted sequential hot deck algorithms, polytomous logistic regression methods, and

multivariate techniques that substantially improve on the standard imputation methods.

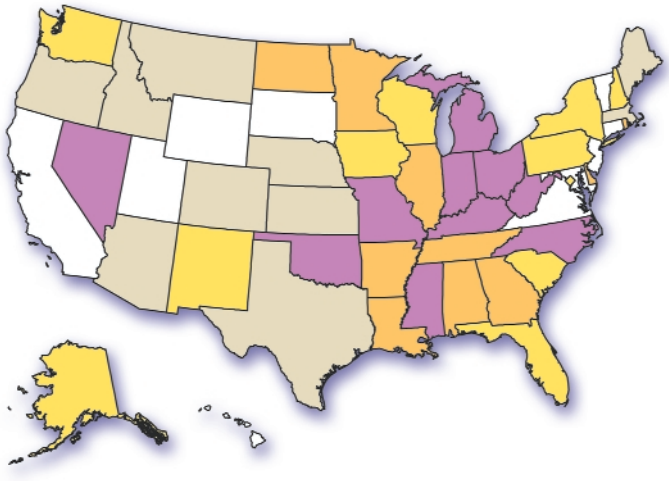
Measurement Error Modeling. RTI statisticians have also made important contributions in response error modeling, including:

- models for studying measurement bias under complex sampling schemes
- multilevel models for describing the effects of interviewers, coders, and other survey personnel on survey response
- probability models for assessing the validity of self-reported drug use
- models for describing the effects of measurement and nonresponse error on categorical data.

Evaluation of Error. RTI statisticians have developed innovative designs and experimental methods for evaluating response error. These range from small-scale laboratory study designs for cognitive research to large-scale field studies to evaluate components of the mean square error or to test survey and study procedures under realistic conditions. RTI leads the field of error evaluation in a number of areas, including re-interview survey design for estimating reliability and response bias, experimental designs for cognitive laboratory research, and designs and statistical methods for comparing alternative modes of data collection.

Small Area Estimation

Small area estimation is one of the newest and most exciting areas of statistical survey research for RTI statisticians. Small area estimation (SAE) is the process of using statistical models to link survey outcome variables to a set of predictor variables known for small domains, in order to predict domain-level estimates. For example, prevalence of a particular disease may be adequately estimated from a study at a national level and it is determined that the prevalence of this disease is highly correlated with a set of predictor variables known at the state level. SAE can be used to produce state-level estimates from this data with far greater precision than one would observe in state-level, direct survey estimates. SAE methodologies are particularly beneficial for policy makers and researchers interested in estimates for small domains but do not have adequate funds for a large-scale survey that could produce precise, direct survey estimates for the small domains.



Application. Additional applications of SAE include estimating for:

- local areas such as neighborhoods using data from national or state studies. Census block and tract level variables, as well as local area statistics like median wage, are often good predictors of many social outcome variables.
- estimating for schools or districts using state level data.
- estimating establishment-level characteristics at a three- or four-digit SIC (Standard Industrial Classification) level using national survey data.

Increased computing power has recently made it possible to use sophisticated hierarchical Bayes methods to fit the mixed logistic and Poisson models that are ideally suited to SAE.

Methodology. RTI statisticians have been developing and implementing SAE methodology for several years. Our most recent work has led us to develop an innovative Survey Weighted Hierarchical Bayes (SWHB) solution for fitting mixed logistic models. This approach has been implemented successfully on data from the National Household Survey on Drug Abuse. Our SWHB solution has several benefits over other hierarchical Bayes solutions obtained from such software packages as WinBUGS® and MLwiN. For example, our SWHB solution makes proper use of the survey weights. Thus, SAEs and posterior predictive density intervals closely match direct survey weighted prevalence estimates and corresponding design-based confidence intervals for areas with large samples and for aggregates of small sample area estimates. RTI's state-of-the-art SAE software is fully capable of producing estimates, for example, for all the block groups (over 226,000) in the United States.

$$\ln\{\pi_{ijk} \div (1-\pi_{ijk}) \mid Z_{aijk} = 1\} = x_{aijk}\beta$$

Biostatistics

Our capabilities in biostatistics cover a broad array of modern methods in the design, conduct, and analysis of studies related to human health. We have served as the data-coordinating center for a variety of multicenter trials in such areas as infectious diseases (including HIV), cardiovascular disease, oncology, pulmonary disease, and obstetrics/gynecology. We conduct national and international epidemiological research into the incidence, prevalence, risk factors, and etiology of diseases. We are leaders in the field of health promotion and disease prevention research and have conducted practical research on community-based programs to promote healthy lifestyle practices and prevent the onset or spread of disease. Our experience includes programs on wellness, preventive care, weight loss, smoking cessation, diabetes care, HIV prevention, drug abuse, and violence reduction.

Biostatistical Expertise. Our expertise includes the design of clinical trials and experimental studies, sample size and power calculations, randomization schemes, and interim stopping rules and analysis plans. We have applied new techniques for analysis of medical and epidemiological studies involving both data that are clustered within clinics and patients, and multiple measurements on the same subjects over time. These analytic procedures, which use generalized estimating equations (GEE), have been integrated into our SUDAAN data analysis system.

Applying Statistical Advances. The breadth of experience and research interests of our statistical staff, and the team-work environment at RTI give our statisticians the very unique ability to bridge the gap between advances in survey design and new research in biostatistics and clinical trials. For example, our expertise in developing multivariate item imputation techniques has direct applicability to clinical trial studies where multiple measurements are taken for each patient. Additionally, our proven techniques using SAS, SUDAAN, and Visual Basic programming to create large numbers of publishable tables and graphs have been successfully utilized by staff to collate efficacy and safety tables for an **electronic NDA submission.**

Statistics for the Behavioral and Social Sciences

RTI has strong capabilities for applying statistical and psychometric techniques to social and behavioral science data.

Statistical Modeling. We are leading the field in the application of generalized linear models such as multinomial logistic regression, Poisson regression, and survival analysis to complex social and behavioral science data sets. We have adapted and applied these models using our SUDAAN software on hierarchical social science data sets where observations within programs, schools, or other units are inter-correlated. We have applied these models on large, complex, national longitudinal survey databases and on smaller and simpler data sets designed to address more narrowly focused issues. Applications have ranged from modeling drug abuse treatment outcomes over time to modeling the impact of special care units in nursing homes on the progression of the symptoms of Alzheimer's disease. We have statisticians formally trained in statistical modeling, and behavioral and social scientists highly experienced in applying these models to their own substantive areas. In addition to new and emerging statistical tools, we frequently use the standard multivariate analysis tools of the behavioral and social scientist (e.g., factor analysis, discriminant analysis, and multivariate analysis of variance).

Psychometrics. We have experts in a broad range of advanced psychometric techniques, including item response theory (IRT), structural equation modeling, cluster analysis, and latent class analysis. We have developed and assessed new scales, using both classical and modern test theory analyses. For example, we have designed studies to assess the reliability, validity, and generalizability of scales. We have also assessed the structure and dimensionality of questionnaires, using confirmatory and exploratory factor analysis, and have related latent scale factors to outcome measures, using structural equation modeling and other latent modeling techniques.

Environmental and Experimental Statistics

RTI's involvement in environmental and experimental statistics encompasses exposure and toxicology studies, as well as laboratory experiments designed to assess the factors that affect the performance of an analytical technique.

Exposure Studies. Since 1980, RTI has been a leader in the design and implementation of studies of human exposures to toxicants. We have conducted studies of airborne pollutants such as volatile organic compounds, polycyclic aromatic hydrocarbons, carbon monoxide, and particles and associated elements such as lead. RTI has also designed studies for estimating ingestion and dermal exposures, as well as ambient environmental (e.g., indoor air, outdoor air, tap water) and body-burden concentrations. We address design issues for environmental research efforts, including ecological studies, residential and nonresidential indoor air studies, and multimedia studies of regional and national scope.

Toxicology Studies. RTI statisticians provide expertise in study design, data analysis, statistical methodology development, software development, and report writing for a variety of toxicology applications for the National Toxicology Program of the National Institute of Environmental Health Sciences (NIEHS), as well as for the pharmaceutical and agrochemical industries. We provide statistical consulting services for a variety of animal and plant toxicology studies conducted under EPA/FIFRA (Federal Insecticide, Fungicide, and Rodenticide Act) good laboratory practices (GLP) guidelines. We are particularly experienced in analyzing safety assessment studies, including neurobehavioral, developmental toxicity and teratology, carcinogenicity, endocrine toxicity, avian toxicity, and acute studies for estimating LD50s and related parameters.

Many of our analyses require specialized statistical techniques. For example,

- Statistical analysis of carcinogenicity data requires statistical methods for evaluating the effect of exposure to a compound on lifetime tumor incidence rates, after accounting for dose-related differences in intercurrent mortality.
- Analysis of neurotoxicology data requires techniques for longitudinal continuous and categorical responses and clustered time-to-event data.
- Statistical analysis of developmental toxicology studies requires techniques that account for intralitter correlation with continuous and binary outcomes.

RTI is also providing experimental design services and quantitative dose-response modeling using linear and non-linear models, as well as analysis of plant toxicology studies (vegetative vigor) and studies of aquatic growth.

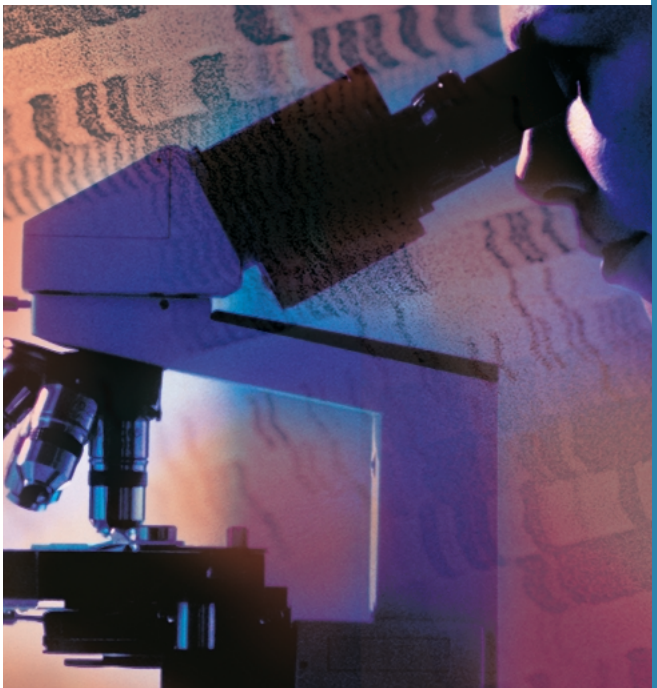
Experimental Design. RTI uses statistical principles of experimental and survey designs in various capacities. For example, we have designed laboratory experiments to assess factors that affect the performance of an analytical technique (e.g., the measurement of cyanide in wastewater). We have also embedded experimental designs in field studies. For instance, we have identified and assessed the magnitude and characteristics of measurement errors and how they vary temporally.

Guidance Documents. RTI statisticians prepare guidance materials for environmental applications. These documents present important concepts in simple language to make the methods understandable to a broad audience. RTI has strong capabilities for applying statistical and psychometric techniques to social and behavioral science data.

$$M_{hi} \quad \Sigma \quad (y_{hij} - \bar{Y}_{hi})^2$$

Bioinformatics—Genomics Research

RTI's newest statistical program area has been developed to provide statistical and software solutions for clients who use genetic and genomic data in their research. The Bioinformatics and Statistical Genetics group within RTI's Statistics Research Division is an interdisciplinary team of professionals with backgrounds in statistics, biomathematics, bioinformatics, and genetics. The group has been working with clients in the biotechnology and pharmaceutical industries and academia, providing consulting and data analysis services in the areas of gene expression analysis, genomic and genetic experimental design, and genetic epidemiology. The group also has extensive experience developing custom software solutions for genetic and genomic research, including the development of databases, visualization software, and automated data analysis pipelines.



For additional information about RTI's SUDAAN and Statistical Consulting Services, visit:

http://www.rti.org/sudaan

or contact:

SUDAAN Statistical Software Center
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709-2194

Phone (919) 541-6602
Fax (919) 541-7431
E-mail: sudaan@rti.org

Open 9:00 AM - 4:00 PM Eastern Time

Research Triangle Institute

Statistics Research Division

turning knowledge into practice