Journal of Official Statistics
Vol. 9. No. 1, 1993, pp. 203–216

Advances in Survey Methods and Technology at the Research Triangle Institute

Paul Biemer, Pat Flynn, Ralph Folsom, Judy Lessler, Robert Mason, James O'Reilly, and Rick Williams[1]

Abstract: This paper provides an overview of the major areas of statistical research at the Research Triangle Institute, with special emphasis on research related to sample surveys. Among the research areas discussed are survey methodology, computer assisted information collection (CASIC), statistical methods for surveys and environmental statistics, and computer software development for data analysis, particularly for complex survey design. The paper concludes with a glimpse of the future of statistical research at the Institute.

Key words: Measurement error; research center; CAPI; data analysis; environmental statistics.

1. Introduction

Research Triangle Institute (RTI) has conducted survey research since 1959, the year RTI was founded. Gertrude Cox, Alva Finkner, and other statisticians led the first survey research projects which were sponsored by the U.S. Department of Commerce and the U.S. Department of Agriculture. Today with a staff of about 1,500 employees and total revenues exceeding $150,000,000 USD a year, RTI is a multidisciplinary organization with programs in many areas in addition to survey research, including: electronics systems; chemistry and life science; social science and international development; analytical and chemical science; environmental science and engineering; biometrics; and statistical science. This diversity greatly benefits the survey research program at RTI which often must draw on this interdisciplinary professional base for subject matter expertise.

In survey research, RTI is perhaps best known for its success in conducting large and complex social surveys. RTI surveys cover a wide range of substantive areas including health, agriculture, education, economics, aging, behavior, and recreation. Recent surveys have included: a national household survey of alcohol and drug abuse; a national survey of 16 to 24 year olds to understand backgrounds, attitudes, motivations and intentions to enlist in the U.S. military; surveys to establish and maintain a national registry of serious diseases, illnesses, and exposure to toxic substances; and a large scale survey to assess the personal health care expenses of the U.S. population. Approximately 90% of the total funding for survey research projects is provided by Federal government agencies.

RTI's first methodological research project

[1] Research Triangle Institute, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, N.C. 27709, U.S.A.

was an evaluation of the 1959 Agricultural Census. Since then, methodological research has played a prominent role in the development of the Institute. This paper describes a number of methodological and technological research areas which are currently active. We focus on major, developing areas; however, many smaller scale methodological research activities are currently underway at RTI in order to meet the growing demand from RTI's clientele for higher quality, lower cost survey research products.

Four research areas have been selected as representative of RTI's current methodological research program. These are: cognitive laboratory research; computer assisted survey information collection (CASIC); statistical methods; and software development for complex survey data analysis. The paper concludes with a glimpse into the future of survey methods and technology.

2. Laboratory for Survey Methods and Measurement

2.1. Overview

One survey enhancement method that is attracting increasing interest from government agencies and research organizations is the examination of thought processes that affect questionnaire answers and response accuracy (Oksenberg and Cannell 1977; Lessler 1987, 1989). Many of these agencies, including the National Center for Health Statistics, the U.S. Bureau of Labor Statistics, and the U.S. Census Bureau, are establishing their own laboratory facilities to investigate these issues. In response to this interest, and in an effort to expand our capability to conduct cognitive research in survey contexts, RTI established the Laboratory for Survey Methods and Measurement in 1986.

The Laboratory for Survey Methods and Measurement is a critical component in RTI's survey research capability. The laboratory contributes to our ability to consistently improve survey measurement through experimental research. The activities conducted in the laboratory both draw on and support other RTI disciplines, such as psychology, clinical assessment, statistics, sociology, gerontology, survey methods, policy analysis, education, economics, and public health. We established the laboratory to systematically research a range of survey methods and measurement issues, such as: questionnaire design; instrument pretesting, revision, and evaluation; response accuracy; procedure effectiveness; instrument validation; public perceptions and attitude assessment; and interviewer assessment. The laboratory, which is located on the RTI campus, is equipped with both permanently installed and portable audio and visual equipment, desktop and notebook computers, and other required laboratory hardware. It is designed to support research in diverse field locations as well as locally.

2.2. Laboratory capabilities

The mission of the laboratory is to enhance the survey methodologies by (1) gathering information on the thought processes that influence the answers repondents give to survey questions, and (2) testing whether the questionnaire is designed to encourage accurate answers. Through these research efforts, important contributions to survey methodology have resulted. Among RTI's contributions are: an evaluation of item count techniques for asking sensitive questions; a question "coding" scheme for analyzing the question-answering demands of survey items; insights into respondents' interpretation of survey items; research on factors affecting respondents' comprehension of reference periods and technical terminology; exploration of specific cognitive demands imposed by questions dealing with substance abuse; effects on responses of prior knowledge, new information content, and source of new information and survey responses; and modeling of the response process inherent in completing self-administered questionnaires.

2.3. *Research methods*

To continue our efforts to improve survey methodology, we use three types of research methods. These methods are:

- *Small Group Research*, including focus group interviews and expert panels. These involve small group research projects involving eight to ten participants who discuss printed or audiovisual materials. The data collected are used to evaluate experimental designs, examine the order of topics in the questionnaires, or investigate respondent comprehension of questions.
- *Survey Interviews*. In the laboratory, face to face interviews are simulated by using printed or audiotaped experimental materials. In addition, telephone lines are available to conduct experimental CATI surveys, and notebook computers are used to simulate field computer assisted interviews either in the laboratory or in the field.
- *In-Depth Cognitive Interviews* to explore how respondents interpret survey questions and how they go about answering survey items. Participants are interviewed individually and sessions are recorded (on audio- or videotape cassettes).

2.4. *Coding systems*

The data collected from research participants are evaluated using a "coding" system for both *cognitive appraisal*, which evaluates the survey instrument, and *behavioral appraisal*, which evaluates the interaction between respondent and interviewer.

2.4.1. Cognitive appraisal

RTI's coding system for cognitive appraisal of survey instruments focuses on issues related to the evaluation of the survey instrument. Through it, the issues of *comprehension*, *interpretation*, *cognitive processing*, *judgment*, and *response* may be examined from the point of view of the respondent. The coding system lists the problems that may arise during an interview. For example, inaccurate responses may be due to the respondent not *comprehending* the question; or, the question may be worded in a way that allows two or more legitimate *interpretations* by the respondent. By examining each of these issues individually for each question, each question as well as the questionnaire as a whole can be improved.

There now are 132 categories in the coding system which is continually being enhanced and refined. Because the coding system is comprehensive, it can be tailored to study the survey instruments for a variety of applications. Some examples of general areas the coding system explores in relation to cognitive appraisal include: ambiguity of technical terminology; confusion between the question asked and the answer given; and hidden instructions for the respondent to do something that is not written.

2.4.2. Behavioral appraisal

RTI also uses a variety of coding systems for behavioral appraisal of survey instruments and procedures (for example, see Cannell and Oksenberg 1988). These coding systems focus on evaluating aspects of the interaction between the interviewer and the respondent. For example, one system targets respondent behavior and evaluates survey items by noting how often respondents (1) attempt to answer items before the interviewer finishes reading them, (2) ask for question clarification, and (3) generate inadequate responses.

These behavior coding results have been used to identify survey items that should be revised, restructured, or clarified.

3. Computer Assisted Survey Information Collection

RTI is conducting computer assisted survey information collection (CASIC) research in a number of areas. Some of these are: computer assisted self-interviewing with optional prerecorded questioning; computer assisted personal interviewing; automated industry and occupation coding concurrent with interviewing; integrated survey processing software; and computer driven field control systems. This section is devoted to a summary of the recent research on the first two items in this list.

3.1. *Audio computer assisted self-interviewing*

RTI is intensively pursuing audio computer assisted self-interviewing (audio CASI) as an alternative to the traditional written self-administered questionnaire (SAQ) for presentation of questions to respondents on drug use, sexual behaviors, and other sensitive matters on in-person surveys.

Audio CASI holds the promise of eliminating two major shortcomings of the written SAQ: (1) the need for literacy on the part of the respondent, and (2) the requirement that questions and logic must be simple and limited, without branching and contingent questions. Also, as developed at RTI, audio CASI offers the prospect of providing a number of significant side benefits, including greater respondent privacy, exactly standardized administration of the questions, and convenient multilingual administration.

With RTI's audio CASI system the computer plays a recorded version of question and answer choices to the respondent over headphones. The subject responds through some data input device (keyboard, external keypad, barcode reader, etc.). The computer records the response and, based on it, plays the next appropriate question.

Supporting functions such as display or blanking of questions on the screen, repeating the question, backing up, and explanations of terms are available immediately when the respondent presses the appropriate key. Other features of computer assisted interviewing such as range and consistency checks and recording the respondent's verbal answers to open-ended items also can be implemented in the audio format.

Technologically, the key element of audio CASI is that recorded questions are stored in digital form on the PC's hard disk and accessible in random fashion. As a result, complex question ordering is available, permitting detailed questioning of those reporting certain behaviors. Earlier audio technologies such as audio tape do not allow for rapid random accessing of items. RTI has developed a working audio CASI system and is engaged in laboratory and field testing. This testing will explore a number of methodological issues related to audio CASI. Among these are:

- Whether audio CASI produces more detailed and reliable responses compared to the written SAQ from illiterates in general, and from other groups for sensitive questions.
- Whether an all-audio or mixed audio–screen mode affects response more substantially.
- How audio CASI affects the pace of the interview, and the ability of respondents to stay focused on the interview.
- What limitations audio CASI holds in

terms of types and lengths of questions and the burden on respondents.

3.2. CAPI

RTI is currently involved in a three-year demonstration project in Washington, D.C., referred to as the DC Initiative, to test the effectiveness of standard and enhanced residential and outpatient drug abuse treatment protocols. In this National Institute on Drug Abuse (NIDA) research demonstration project, RTI is using an automated data collection system and computer assisted personal interviews (CAPI) to conduct face to face client interviews. The primary data collection instrument for this project is the Individual Assessment Profile (IAP) which is a structured interview instrument designed for use with substance-abusing populations in several large-scale national treatment projects. One of the primary reasons behind our decision to use CAPI is that it provides immediate data, feedback, and information through the automated reporting functions. We believe this will increase the overall quality of the data.

RTI is also collaborating with the National Opinion Research Center (NORC) on the National Treatment Improvement Evaluation Study (NTIES) which was designed to evaluate a multi-million dollar Federal program to improve drug abuse treatment. The primary purpose of NTIES is to determine whether enhanced or more comprehensive treatment, which is provided through government funded demonstration programs, is more effective than the less comprehensive treatment which existed prior to the initiative and infusion of Federal dollars.

Staff in treatment centers have concerns as to the effect of CAPI on the respondents' willingness to report sensitive information. Therefore, in the early phase of the NTIES evaluation, RTI will be testing the feasibility of CAPI for possible use throughout the entire evaluation. We will assign treatment programs to be evaluated to one of two groups: (1) CAPI, where a modified version of the Individual Assessment Profile (IAP) will be administered through the use of laptop computers, and (2) traditional interview schedule, where the modified IAP will be administered through the use of traditional face to face, paper-and-pencil survey interviewing methods. After approximately three months, the data from both groups will be analyzed to identify response differences to sensitive items as well as socially desirable and undesirable response categories.

A third major evaluation study in which RTI will be employing CAPI methodology is in the Evaluation of Campus Treatment Demonstration Programs. The purposes of the Campus Treatment Demonstration Programs are to increase residential drug treatment capacity, demonstrate useful residential treatment models, and evaluate the efficacy and efficiency of campus-based approaches to residential drug treatment. RTI staff will conduct clinical trials which compare and contrast various treatment protocols to determine their efficacy. Research and clinical data will be obtained from patients through the use of the CAPI–IAP at intake and at critical intervals during the course of treatment. One of the primary reasons behind the use of CAPI is to provide immediate data, feedback, and information through the automated reporting functions that were developed along with the CAPI–IAP application. As in the DC Initiative research project, client narrative reports will be generated through the CAPI system to provide immediate clinical feedback for treatment planning and for compliance with State and Federal reporting requirements.

In addition to our research on response to sensitive questions, we will conduct several additional empirical studies of the CAPI–IAP. The primary goal of these proposed investigations is to determine the reliability of the IAP when administered by program counselors/clinicians versus professional program researchers and to determine the IAP's reliability when administered by CAPI versus traditional personal interviewing.

4. Statistical Methods

4.1. Probability sampling and estimation at RTI

RTI has a large group of sampling statisticians who conduct research in survey methodology, sample design, selection, estimation, and evaluation. Many of the statistical methods now in use for surveys at RTI are products of RTI's sampling research program. Before describing the activities currently underway in the program, the history of survey statistics at RTI from the past two decades will be briefly reviewed, mentioning some of the most notable contributions in this area.

In the sample design and selection area, two developments have strongly influenced RTI's approach. Chromy's sequential probability minimum replacement (PMR) sampling algorithm has become our principal mode of selection for complex probability samples (Chromy 1979, 1981; Williams and Chromy 1980). Chromy's method has all the appealing features of a PPS systematic sample including the implicit stratification effect of a judiciously ordered frame. It has the advantage of unbiased variance estimability. Our approach to formal sample design optimization has been strongly influenced by Chromy's algorithm for minimizing linear survey cost models subject to multiple variance constraints (Chromy 1987). This algorithm handles a wide variety of sample design optimization problems including optimum sample allocations to design strata and optimum cluster size determinations. The provision for multiple variance constraints allows one to explicitly balance the competing precision demands of a multipurpose survey.

Another important sample design and selection tool that has been refined at RTI is the "composite size measure" methodology (Folsom, Potter, and Williams 1987). This methodology facilitates the oversampling of target subpopulations in clustered samples. Use of a composite size measure extends the appealing features of PPS cluster sampling to the case of multiple subpopulation oversampling; namely, equal cluster sizes are maintained at each stage of selection while uniformly achieving the targeted selection probabilities within subpopulations. This methodology has proven to be particularly advantageous for large national surveys where face to face interviewing is combined with requirements for over-representing particular demographic subgroups.

Turning to the area of survey estimation, RTI has been instrumental in the development of efficient computer software for the appropriate analysis of data derived from probability samples. While this software development is discussed in Section 5, it should be noted that some of the pioneering work in the derivation of efficient δ-method or Taylor series variance estimators for complex survey statistics was done at RTI (Folsom 1974). Folsom derived the analytic form for the linearized variate that was subsequently used to compute the variance–covariance matrix estimator in RTI's survey regression procedure SURGGER (cf. Shah, Holt, and Folsom 1978).

More recently, Folsom (1984, 1986) and Williams (1989) have developed a probability sampling theory for Hoeffding's (1948) U-statistics. The U-statistics are generalized symmetric means that are used in classical nonparametric inference for variance components, measures of association, and various nonparametric rank tests. For survey statistics, the nonparametric variance component and measure of association applications are apparent. Other survey applications of U-statistics include robust estimators for the degrees-of-freedom parameters in Student's t and Snedecor's F type statistics that incorporate sample design-based variance estimators.

RTI has also developed probability sample based modes of estimation and inference for longitudinal or panel survey data. Chambless and Boyle (1985) applied Binder's (1983) extension of δ-method variance estimation for implicitly defined statistics to the discrete time proportional hazards model. Folsom, LaVange, and Williams (1986) discuss several panel data analysis topics from a probability sampling perspective. These topics included longitudinal family definition and weighting, general linear modeling of growth curves, Kaplan-Myer product moment estimators, and the continuous time Cox proportional hazards model (Binder 1992).

We have also developed improved methods for dealing with survey nonresponse and missing questionnaire data. Folsom (1991), and Iannacchione, Milne, and Folsom (1991) present new weight adjustment methods based on generalized exponential and logistic raking. The connection between our generalized raking weight adjustment estimators and linear regression/imputation estimators is established. Simple δ-method variance approximations employing linear regression residuals are derived. Potter's (1990) recent work on weight trimming explores the relative merits of several methods for reducing the adverse variance inflation effects of extreme survey analysis weights. In the area of missing questionnaire data imputation, Cox's (1980) weighted sequential hot deck algorithm improves on the standard unweighted sequential hot deck in several respects. Within imputation classes, it preserves the weighted respondent data distribution in expectation over repeated stochastic imputations. It also minimizes the number of times an imputation donor is used. Cox's hot deck algorithm exercises additional control over imputation bias and variance within classes by sequentially selecting imputation donors from a purposively ordered list and matching them sequentially to missing data cases sorted in the same order.

Survey response error modeling is another area where RTI statisticians have made important contributions. Former Executive Vice-President Dan Horvitz's leadership has strongly influenced these efforts. Lessler's (1976) early work on double sampling difference estimators is a notable example. Horvitz's advocacy of total survey error modeling (Horvitz 1980; Horvitz and Wolter 1975) was the inspiration for Lessler and Kalsbeek's work that has culminated in Lessler and Kalsbeek (1992). Other collaborations in this area include the modeling of and adjustment for response bias in the U.S. National Crime Survey (LaVange and Folsom 1985) and the use of standards in survey design (Folsom, Horvitz, and LaVange 1989) and in survey estimation (Horvitz, Folsom, and LaVange 1989).

4.2. *Environmental statistics*

Although there is some confusion as to the exact meaning of the term "environmental statistics," RTI's contributions in this general area began in the late 1970s. Since that time we have been heavily involved in designing and conducting surveys of: chemical substances in surface water, ground water, drinking water, soils, household dust, indoor and ambient air; biological indicators of surface water quality; human exposures, both occupational and nonoccupational, to pollutants and contaminants; human tissue and excretion burdens of various chemicals; business establishments, such as the water supply industry, hazardous waste generators, commercial incinerators, and establishments with underground storage tanks; household pesticide use; and domestic energy use, and in studying methodological issues arising in association with these types of surveys. The work has been sponsored by state and Federal government agencies and by private sector clients. Most of the surveys have been fully national in scope, although some were restricted to areas defined by the production of specified agricultural crops or sales of specified agrochemicals.

The statistical and methodological issues arising in association with these types of surveys differ in degree but not in principle from those usually encountered in survey research. For example, the existence of an inferential population consisting of distinct and countable units seldom presents a conceptual problem in survey research in general. However for many, if not most, environmental surveys, "natural" units, arising in a self-evident way given the objectives of the study, are unlikely to exist. What are the choices of units, for example, for studying surface water, ground water, ambient air, hazardous waste sites, and underground storage tanks? Complication is added by the fact that the parameters of inferential interest to environmental surveys are most likely to be temporally varying quantities, such that the population units are defined in both time and space, requiring partitioning of the time continuum into distinct, but more or less arbitrary, time intervals. The choices of population units are obviously bounded below by the finite amount of material and the finite length of time needed to collect the observations or make the measurements. Practical considerations, however, are likely to dictate more grossly defined lower bounds.

Often the choice of units evolves as a compromise among statistical, operational and measurement considerations, with most of the emphasis placed on measurement. The units must be of such a size and duration that the procedures used to obtain the measurements will yield values that have acceptable bias, given the study objectives, and variances that are small in relation to the variability that exists among the units themselves. Riverine surface water populations can be constructed using reaches (the water volumes between the confluences of a river or stream) defined at a particular map scale. A scale of 1 : 50,000 has been used for some biological measures (Galuz 1984), whereas a more detailed scale, showing smaller reaches on average, might be required for chemical measurements (Mason 1982, personal correspondence). In any case, the measurement procedures employed need to integrate the variability that exists over the reach length, width, depth, flow, bottom type, etc., such that the measurement obtained accurately characterizes the entire reach during a specifiable time interval.

The problem of unit definition carries over to sampling frame construction. Frame construction is viewed as a process whereby each unit in the population is identified and access is provided to the units for the dual purposes of selecting the sample and obtaining the measurements. Acquiring the source information for frame construction, including information for computing size measures and defining strata, may itself require an extensive data collection operation over several stages of sampling. For example, strata based on successively more localized information about crop distributions and ground water vulnerability to contamination have been used in multistage surveys of pesticide residues in domestic wells (Whitmore et al. 1989; Mason et al. 1988). Strata constructed to provide investigator control over data collection costs assume an inflated importance in many environmental surveys, given per unit data collection costs that may be several thousands of dollars and the often remote and isolated locations of some elements of the sample. A complication is added if the stratification variables themselves are temporally varying quantities. In sampling estuaries, for example, local stratification may be required with respect to seasonally and possibly diurnally varying density gradients, polyhaline and oligohaline boundaries (Mason 1981).

Regardless of how units may be defined, the available measurement technology may act to render many, sometimes most, of the population units non-observable. For example, pesticide residue concentrations that are below the detection limit of the best chemical analytic procedure available for use in a survey obviously cannot be measured. If the pesticide in question is but sparsely distributed, response variable values cannot be obtained for most, perhaps all, of the units in the sample. Non-observability in this sense implies that some parameters, population means for example, are not estimable. For many investigations, a satisfactory way around the non-observability problem is to define the parameters of interest in terms of the proportions of the population that exhibit specified concentrations of a chemical, with "below the detection limit" being the lowest tabulated concentration. In effect the set of proportions estimates a discrete density function that can approximate the distribution of chemical concentrations to an arbitrary degree. Design issues then center on determining the sample size and allocation required to provide specified probabilities of observing or detecting given population proportions (Mason and Lucas 1990, personal correspondence 1990).

Although the concept of a method detection limit is clear, its construct historically has been inconsistently formulated. RTI research has included the development of a rigorous definition of the concept and experimental procedures for its estimation (Clayton, Hines, Hartwell, and Burrows 1986). Furthermore, RTI has contributed to the area of density estimation in the presence of measurement errors (Folsom 1987).

5. Computer Software for Survey Data Analysis

RTI has a long history of developing and maintaining state of the art software for the design-based analysis of survey data. In the 1970s, RTI first developed its STDERR procedure. This first general use package estimated means, totals, proportions and their design-based variance estimates for user-specified subgroups for a variety of stratified, unequally weighted, multi-stage sample designs. This was soon followd by the SURRGER procedure, which added the estimation of linear regression models. During the 1980s, additional data analysis procedures (RTIFREQS, SESUDAAN, RATIOEST, and RTILOGIT) were added to include cross-tabulations, standardized subgroup comparisons, estimates of general ratios ($\Sigma y_i / \Sigma x_i$), and logistic regression models. All of the above procedures were implemented as supplemental SAS (copyright SAS Institute, Inc.) version 5 procedures running on IBM mainframe computers under MVS operating system.

RTI has consolidated and revamped all of these procedures into the SUDAAN (SUrvey DAta ANalysis) system. With support from the U.S. Public Health Service, SUDAAN was developed as an integrated survey data analysis system written in the C language as transportable software not specific to any equipment or operating system. SUDAAN currently runs on IBM compatible personal computer (MS–DOS operating system), DEC VAX mainframes and workstations (VMS or ULTRIX operating systems), Sun SPARCstations (SunOS operating system), and IBM mainframes (MVS operating system). SUDAAN uses a SAS like user interface, which is familiar to most statistical analysts. In addition, SUDAAN can directly access either SAS data sets or ASCII files.

Currently, SUDAAN offers three descriptive and two modeling procedures. The descriptive procedures (CROSSTAB, RATIO, and DESCRIPT) produce estimates of means, totals, proportions, cross-tabulations, generalized ratios, geometric means, quantiles and their estimated standard errors. The descriptive procedures can also produce chi-square tests for independence in two-way contingency tables; standardized comparisons; post-stratified estimates; and contrasts among domain estimates. The two modeling procedures (REGRESS and LOGISTIC) estimate linear regression models for continuous outcome variables and logistic regression models for binary outcome variables. Both procedures allow continuous and categorical independent variables and produce hypothesis tests concerning the independent variables. Multiple degree of freedom linear contrasts of the coefficients can be specified and tested. Finally, coefficient estimates and their covariance matrix may be saved for additional analysis.

Variance estimation for all statistics is based on the first-order Taylor series method of linearization. Multiple design options allow SUDAAN to analyze data from many different complex sample designs. Users may specify variance estimation for (1) sampling with replacement at the first stage of selection (with equal or unequal probabilities); (2) simple random sampling at each stage (with or without replacement); and (3) unequal probability sampling without replacement at the first stage and simple random sampling (with or without replacement) at the subsequent stages. All three options allow for stratification. SUDAAN's options cover a wide range of the most commonly used stratified, multi-stage designs selected with unequal probabilities.

RTI is currently developing two more SUDAAN procedures. A general categorical data analysis procedure (CATAN) will offer log-linear modeling of a contingency table. The SURVIVAL procedure will offer both discrete and continuous proportional hazards models for survival time data.

RTI is exploring the use of survey data analysis methods and software in other areas of data analysis where nested error structures are found. For example, we are exploring the use of SUDAAN as a method for analyzing data from teratology studies. The typical teratology screening experiment involves administration of a compound to pregnant dams of a rodent species, followed by evaluation of the fetuses for various types of malformations. The intralitter correlation in teratology studies is equivalent to the intracluster correlation found in multistage sample designs. The variance estimation methods implemented in SUDAAN for multi-stage sample designs apply directly to the clustered (by litter) teratology data. Our initial simulation results suggest that SUDAAN will provide a valuable new data analysis option for teratology studies.

6. Future Directions

The field of survey research has been tremendously affected by recent technological and methodological innovations. Some of these are: CATI; CAPI; cognitive research methods; the total design method (TDM) for mail surveys; computer assisted data entry (CADE); and user-friendly survey data analysis software packages. Other technological innovations which could substantially affect the field in the future include: audio CASI; computer "pad" or "pen" data collection; touch-tone data entry; and videotelephonics. These advances, which promise to revolutionize the way researchers collect survey data in the future, have survey methodologists racing to keep up with the changes and to become knowledgeable about their effects on survey response. In the future, considerable effort must be devoted to the systematic investigation of cost and quality properties of modern data collection methods so that their use will be thoughtfully guided.

In addition to these technological breakthroughs, survey research in the U.S.A. will be affected by continual quality improvement, following the current trends of American industry. We expect that survey sponsors and users will continue to be more demanding of quantifiably accurate and cost effective data. In this regard, greater emphasis will be placed on quality measures (such as coding error rates, response variance components, edit failure rates, etc.) associated with all aspects of the survey process.

In the future, we anticipate that RTI will integrate its laboratory work with statistical estimation in order that the laboratory work will lead to a more complete understanding of the accuracy and reliability of alternative survey measurement methods. More accurate measurement methods will undoubtedly be more costly to implement thus leading to surveys that make use of embedded experiments, i.e., experiments in which alternative measurement methods are assigned to subsamples on the basis of the cost and the accuracy of the different methods. The estimation of population parameters will subsequently be based upon models – regression models, memory decay models, and so on – that use the information gathered during the field survey/experiment. Concomitantly, we anticipate that RTI's survey analysis software will be expanded to permit routine analysis of surveys with embedded experiments.

This software expansion will include robust probability sampling methods of estimation and inference for hierarchical or mixed (fixed and random effect) models. High priority will be given to the development of a logistic regression procedure that includes Gaussian random effects for natural population and survey defined clusters. Work will also continue on our exponential and logistic raking procedures. Our immediate goal is software that will calculate the associated δ-method linearized variates and standard errors for raking adjusted subpopulation proportions and means. We envision further advances in our statistical methods/software for dealing with questionnaire item nonresponse. To this end, we are exploring simultaneous estimation of logistic regression models for (1) the propensity to respond to a questionnaire item, and (2) the probability distribution of the associated categorical questionnaire item.

An ever-increasing need for high quality, low cost survey data by government agencies and commercial organizations guarantees a strong future for social surveys and the survey organizations who conduct them. However, free-market competition and tighter Federal budgets will exert pressure on American survey organizations to seek better methods and technologies for producing survey information. Thus, the goals of RTI's survey methods research and development program are aimed at meeting these future challenges.

7. References

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz,N.A., and Sudman, S. (Eds.) (1991). Measurement Errors in Surveys. New York: John Wiley.

Binder, D.A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. International Statistical Review, 51, 279–292.

Binder, D.A. (1992). Fitting Cox's Proportional Hazards Model from Survey Data. Biometrica, 79, 1.

Cannell, C.F. and Oksenberg, L. (1988). Observation of Behavior in Telephone Interviews. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls, and J. Waksberg, Telephone Survey Methodology, New York: John Wiley, 475–495.

Chambless, L.E. and Boyle, K.E. (1985). Maximum Likelihood Methods for Complex Sample Data: Logistic Regression and Discrete Proportional Hazards Models. Communications in Statistics – Theory and Methods, 15, 1377–1392.

Chromy, J.R. (1979). Sequential Sample Selection Methods. Proceedings of the Survey Research Section, American Statistical Association, 401–406.

Chromy, J.R. (1981). Variance Estimators for a Sequential Sample Selection Procedure. In D. Krewski, R. Platek, and J.N.K. Rao, (Eds.), Current Topics in Survey Sampling, New York: Academic Press, 329–347.

Chromy, J.R. (1987). Design Optimization with Multiple Objectives. Proceedings of the Section on Survey Research Methods, American Statistical Association, 194–199.

Clayton, C.A., Hines, J.W., Hartwell, T.D., and Burrows, P.M. (1986). Demonstration of a Technique for Estimating Detection Limits with Specified Assurance Probabilities. Research Triangle Institute, RT1/2757/05–01F.

Cox, B.G. (1980). The Weighted Sequential Hot Deck Imputation Procedure, Proceedings of the Section on Survey Research Methods, American Statistical Association, 721–726.

Folsom, R.E. (1974). National Assessment Approach to Sampling Error Estimation: Sampling Error Monograph. RTI final report prepared for the Education Commission of the States', National Assessment Project.

Folsom, R.E. (1984). Probability Sample *U*-Statistics: Theory and Applications for Complex Sample Designs. Institute of Statistics Mimeo Series No. 1464, University of North Carolina, Chapel Hill, N.C.

Folsom, R.E. (1986). Probability Sample *U*-Statistics: Theory and Applications for Complex Samples. Proceedings of the Section on Survey Research Methods, American Statistical Association, 75–80.

Folsom, R.E. (1987). Estimating the Uncontaminated Radon Distribution. Research Triangle Institute, RT1/7810/02–03F.

Folsom, R.E. (1991). Exponential and Logistic Weight Adjustments for Sampling and Nonresponse Error Reduction. Proceedings of the Social Statistics Section, American Statistical Association, in press.

Folsom, R.E., Horvitz, D.G., and LaVange, L.M. (1989). The Design of Surveys Using Measurement Design Standards. Proceedings of the Section on Survey Research Methods, American Statistical Association, 80–87.

Folsom, R.E., Potter, F.J., and Williams, S.R. (1987). Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains. Proceedings of the Section on Survey Research Methods, American Statistical Association, 792–796.

Folsom, R.E., LaVange, L.M., and Williams, R.L. (1986). A Probability Sampling Perspective on Panel Data Analysis. In D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh (Eds.), Panel Surveys, New York: John Wiley, 108–138.

Glauz, W.D. (1984). 1982 National Fisheries Survey, Volume II: Survey Design. U.S. Environmental Protection Agency and U.S. Dept. Interior, FWS/OBS–84/14.

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. Annals of Mathematical Statistics, 19, 293.

Horvitz, D.G. and Wolter, K. (1975). Total Survey Design. NCHSR Research Proceedings.

Horvitz, D.G. (1980). On the Significance of a Survey Design Information System. Proceedings of the Section on Survey Research Methods, American Statistical Association, 122–124.

Horvitz, D.G., Folsom, R.E., and LaVange, L.M. (1987). The Use of Standards in Survey Estimation. Proceedings of the Section on Survey Research Methods, American Statistical Association, 546–551.

Iannacchione, V.G., Milne, J.G., and Folsom, R.E. (1991). Response Probability Weight Adjustments Using Logistic Regression. Proceedings of the Section on Survey Research Methods, American Statistical Association, in press.

LaVange, L.M. and Folsom, R.E. (1985). Regression Estimates of National Crime Survey Operation Effects: Adjustment for Non-Sampling Bias. Proceedings of the Social Statistics Section, American Statistical Association, 109–114.

Lessler, J.T. (1976). Survey Designs Which Employ Double Sampling Schemes for Elimination Measurement Process Bias. Proceedings of the Social Statistics Section, American Statistical Association, 520–525.

Lessler, J.T. (1987). Use of Laboratory Methods and Cognitive Science for the Design and Testing of Questionnaires. Paper prepared for Statistics Sweden, R&D Report, Statistics Sweden, U/STM–35.

Lessler, J.T. (1989). Reduction of Memory Errors in Survey Research: A Research Agenda. Paper presented to International Statistical Institute, Paris, 1989.

Lessler, J.T. and Kalsbeek, W. (1992). Nonsampling Error in Surveys. New York: John Wiley.

Mason, R.E. (1981). A Statistical Design for a National Estuary Monitoring Network. Research Triangle Institute, RTI.1864/14–04F.

Mason, R.E., Piper, L.L., Alexander, W.J., Pratt, R.W., Liddle, S.K., Lessler, J.T., Ganley, M.C., Munch, D.J., and Langner, G. (1988). National Pesticide Survey Pilot Evaluation. Technical Report, Research Triangle Institute, RTI/7801/06–02F.

Oksenberg, L. and Cannell, C.F. (1977). Some Factors Underlying the Validity of Response in Self-report. Bulletin of the International Statistical Institute, 325–346.

Potter, F.J. (1990). A Study of Procedures to Identify and Trim Extreme Sampling Weights. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Shah, B., Holt, M.M., and Folsom, R.E. (1978). Inference About Regression Models for Sample Survey Data. Bulletin of the International Statistical Institute, XLVII, 43–57.

Whitmore, R.W., Pratt, R.W., Piper, L.L., Rush, G.A., Alexander, J.W., Liddle, S.K., Ganley, M.C., and Truesdale, R.S.

(1989). National Alachlor Well Water Survey, Vol. I: Survey Design and Data Collection. Research Triangle Institute, RTI/3895/04–02D.

Williams, R.L. (1989). Large Sample Theory for *U*-Statistics in Unequal Probability Samples. Ph.D. Dissertation, Department of Biostatics, University of North Carolina at Chapel Hill. Institute of Statistics, Mimeo Series No. 1860T.

Williams, R. and Chromy, J.R. (1980). 'SAS' Sample Selection Macros. Proceedings of the 1980 SAS Users Group International.

Received March 1992