An Evaluation of Alternative PC-Based Software Packages Developed for the Analysis of Complex Survey Data

Steven B. COHEN

Data from complex survey designs require special consideration with regard to variance and data analysis as a consequence of significant departures from simple random sampling assumptions. Using data from the National Medical Expenditure Survey (NMES), which is characterized by a complex survey design, three programs developed for the analysis of complex survey data in a personal computing environment are compared: Stata, SUDAAN, and WesVarPC. The comparisons concentrate on user facility, computational efficiency, and program capabilities.

KEY WORDS: Complex survey data; NMES; Stata; SUDAAN; WesVarPC.

1. INTRODUCTION

National surveys conducted by government organizations, industry, political organizations, and market research firms often share the same survey design objective to minimize the variance in survey estimates, subject to fixed cost and time constraints. As a consequence, most large-scale national health care surveys are characterized by sample designs with varying degrees of complexity, with design features that include clustering, stratification, disproportionate sampling, and multiple stages of sample selection. Most of the standard statistical software packages such as SAS, SPSS, SYSTAT, and BMDP assume that the data were obtained from a simple random sample in which the observations are independent and identically distributed, and selected with equal probability. When the data have been collected from a survey with a complex sample design, variance estimates of survey statistics derived under simple random sampling assumptions generally underestimate the true variance, which results in artificially lower confidence intervals and anticonservative hypothesis testing, that is, rejecting the null hypothesis when it is true, more frequently than indicated by the nominal Type I error level (Carlson, Johnson, and Cohen 1993).

Steven B. Cohen is Director, Statistical Research and Methods Division, Center for Cost and Financing Studies, Agency for Health Care Policy and Research, Rockville, MD 20852 (E-mail: scohen@cghsir.ahcpr.gov). The views expressed in this paper are those of the author, and no official endorsement by the Department of Health and Human Services or the Agency for Health Care Policy and Research is intended or should be inferred. The author especially thanks Mamatha Pancholi of the Center for Cost and Financing Studies, AHCPR, for her statistical assistance, Steve Machlin, AHCPR, for his detailed review, Dr. Sam Zuvekas, AHCPR, for his assistance with the use of the Stata package, and Dave Naden, SSS, for his data management support. Each of the developers of the statistical packages under review was most helpful in providing materials and answering technical questions. Special thanks to Dr. Richard Goldstein and the reviewers for their insightful comments.

In the past decade a number of statistical software packages have been developed, specifically tailored to facilitate the analysis of complex survey data (Cohen, Burt, and Jones 1986). This evaluation is directed to three software packages that have been developed to facilitate the analysis of complex survey data in a personal computing environment. The software packages under review are: Stata release 5.0 with the survey design software added, SUDAAN Version 7.0, and WesVarPC Version 2.02. Data from the household component of the 1987 National Medical Expenditure Survey (NMES), which has a multistage complex survey design, was used to facilitate the evaluation. The comparisons focus on analytical capacity, programming ease, computer run time, documentation, and data preparation issues.

2. BACKGROUND

A number of alternative methods have been developed for approximating sampling variances for survey estimates derived from surveys with complex sample designs. Three generally accepted and frequently used techniques are the Taylor Series linearization method, the method of balanced repeated replication (BRR), and the jackknife method (Wolter 1985). A number of prior software evaluations have focused on software packages developed for mainframe computing applications (Cohen et al. 1986; Cohen, Xanthopolous, and Jones 1988). With the enhanced computing capacity made available on the PC, more attention has been given to the development of software packages tailored to the analysis of complex survey data in a PC environment. In addition to the packages that are part of this evaluation, another software package developed for the analysis of complex survey data in a personal computing environment, PC CARP, is available. The advantages and disadvantages of this software procedure relative to an earlier version of SUDAAN were considered in a prior evaluation. Because no upgrades have been implemented to the PC CARP statistical package at the time of this evaluation, the readers are directed to the results of that evaluation (Carlson et al. 1993). It should be noted that SAS, SPSS, and Systat were also invited to be part of the evaluation in order to identify new capabilities with respect to the analysis of complex survey data, but they declined.

To conduct the software comparison, data from the household component of the 1987 National Medical Expenditure Survey were used. The household component of the 1987 National Medical Expenditure Survey (NMES) was designed to produce unbiased national and regional estimates of the health care utilization, expenditures, sources of payment, and health insurance coverage of the U.S. civilian noninstitutionalized population. The survey was sponsored by the Agency for Health Care Policy and Research. The NMES household survey was a year-long panel that collected measures of health status, use of health care services, expenditures and sources of payment, insurance coverage, employment, income and assets, as well as demographic information for calendar year 1987. The survey adopted a stratified multistage area probability sample design, consisting of 165 primary sampling units, 2,317 area segments, and about 15,000 responding households, which represented the union of two independently drawn national samples of households, one by Westat, Inc. and one by NORC, the National Opinion Research Corporation, the NMES data collection organizations. The primary sampling units were defined as counties or groups of contiguous counties, and the segments were defined as census blocks or enumeration districts. The combined sample of 165 primary sampling units (81 from the Westat sample; 84 from the NORC sample) consisted of 127 distinct sites. The NMES design required selective oversampling of blacks, Hispanics, the poor and near poor, those 65 years of age and older, and the functionally limited or impaired. An initial screening interview was conducted in the fall of 1986 for a sample of approximately 35,600 addresses to obtain information required for oversampling. A subsample of about 15,000 households was selected for the detailed interviews. Field operations for the NMES household component consisted of four core interviews conducted with the selected households at three–four month intervals over a 15-month period. The first two interviews were conducted in person, the third by telephone (if a telephone was available and the procedure was acceptable to the household), and the final core interview in person (Edwards and Berlin 1989). The overall survey response rate was 80.1%, and yielded 34,459 individuals with complete information for 1987 for their entire period of survey eligibility (Cohen, DiGaetano, and Waksberg 1991).

The NMES Research Findings Series consists of reports presenting national estimates of several key measures of health insurance coverage, use of and expenditures for inpatient hospital services, outpatient visits, ambulatory office-based physician visits, dental services, home health visits, and prescribed medicines, in addition to their sources of payment. The data analyses necessary to create these reports follows a similar pattern: 1) data processing to create relevant analytical files and to derive weighted national estimates of the specified health care measures, 2) generation of standard errors associated with each health care estimate presented in the NMES reports, and 3) performance of hypothesis tests to compare estimates across population subgroups.

Similar procedures were undertaken to produce the NMES health care estimates, their related standard errors, and the conduct of hypotheses tests for over 100 AHCPR analytical papers using the NMES data. As a consequence of the frequency of generating the standard errors for NMES reports, a study of the capabilities and efficiency of alternative statistical software procedures developed for PC applications would have implications regarding the software procedures to adopt for the generation of reports using data from the current 1996 National Medical Expenditure Panel Survey (MEPS). The MEPS is the most recent national medical expenditure survey, and is cosponsored by the Agency for Health Care Policy and Research and the National Center for Health Statistics.

Many of the survey design complexities inherent in the NMES design are comparable to other national health care surveys such as the National Health Interview Survey, the Medicare Current Beneficiary Survey, and the National Health and Nutrition Examination Survey (Cohen 1996). Furthermore, several of the analytical requirements of the NMES survey are generally convergent with other national health care surveys in the derivation of descriptive statistics and the conduct of multivariate analyses. Consequently, a comparison of the performance of the alternative statistical software packages in their capacity to generate national estimates of health care parameters from the NMES data that are expressed in terms of population means, totals, and ratio estimates and the estimation of associated standard errors and related design effects should be relevant to other national survey efforts with similar design features and analytical requirements.

2.1 Study Design

To facilitate the software comparisons tables presenting health care estimates in the same format as the NMES Research Findings Series were specified. Three distinct sets of statistics were identified, presenting health care estimates in terms of means, totals, and ratio form. For statistics expressed in terms of means a table presenting mean overall health care expenditures (TOTALEXP), mean out of pocket health care expenditures (TOTALSP1), and mean health care payments by private insurance carriers (TOTALSP2) was specified (Table 1). In this table the mean expenditures and source of payments were derived for the overall population, and also were restricted to the population with an expense. The inclusion of a set of estimates that were restricted to a subset of that database (e.g., individuals with health care expenses) also provided an opportunity to consider the effect of sample size on computational efficiency. In general, standard estimates for subpopulations should be derived using the entire dataset of interest, and specifying a subpopulation command in a given software package. In addition to producing expenditure estimates for the nation, the more detailed table contained mean estimates of the three criterion variables for domains defined by the mutually exclusive and exhaustive categories of eight demographic measures. They included age (AGE: <6, 6–17, 18–44, 45–64, 65+), sex (SMPSEXR), race/ethnicity (RACE3: Hispanic, black–non-Hispanic, white–non-Hispanic, and other), family income classification relative to poverty level (POVSTAL: poor, near poor, low income, middle income, high income), size of residential area (SMSA: 19 largest MSA, other MSA, non-MSA) census region (SREGION), census division (CENDIV), and perceived health status (RATEHLTH: excellent, good, fair, poor). For comparative purposes the following set of additional demographic breakdowns was considered: cross classifications of the respective poverty status and health status measures, and all two-way and three-way cross classifications of age, race/ethnicity, and gender. The estimates presented in Table 1 are condensed versions of the full set of estimates that were considered in the evaluation and would appear in a standard table in the NMES Research Findings Series. A table presenting statistics in terms of totals was also examined (table and

Table 1. Expenditures and Sources of Payment for Health Care Services: Mean Expense and Source of Payment, United States, 1987

Population characteristic	Population in 1,000s	Mean expense (s.e.)	Mean out-of-pocket payment (s.e.)	Mean payment by private insurance (s.e.)	Mean expense (s.e.)	Mean out-of-pocket expense (s.e.)	Mean payment by private insurance (s.e.)
		Per person	Per person	Per person	Per person with expense*	Per person with expense	Per person with expense
Total	239,393	$1520.85 (40.26)	$359.71 (8.97)	$550.93 (17.55)	$1800.87 (46.46)	$425.93 (10.39)	$652.37 (20.54)

*202,170,000 individuals with health care expense.
Source: 1987 NMES Household Survey, AHCPR.

estimates not shown) for purposes of comparison for the same health care measures, using the same format with respect to the domains defined by the mutually exclusive and exhaustive categories of eight demographic measures, cross classifications of the respective poverty status and health status measures, and all two-way and three-way cross classifications of age, race/ethnicity, and gender.

An additional table that focused on national estimates of sources of payment information for use of health care services was specified in ratio form (Table 2). Six mutually exclusive and exhaustive categories regarding sources of payment associated with overall health care expenditures for health care services were specified: amount paid out of pocket by family, amount paid by private insurance, amount paid by Medicare, amount paid by Medicaid, amount paid by other public programs, and amount paid by other payers. For each of the six categories a ratio estimate was defined as the ratio of the amount paid by source of payment to the total medical expenditures for all health care services utilized in calendar year 1987. As before, the table format mirrored the format for the full set of estimates that were considered in the evaluation and would appear in a standard table in the NMES Research Findings Series with respect to the demographic breakdowns. Again, a table presenting statistics in terms of totals was also included for purposes of comparison for the same health care measures, using the same format with respect to the demographic breakdowns (table and estimates not shown).

3. SOFTWARE COMPARISONS

The three software procedures, Stata, SUDAAN, and WesVarPC, were then compared with respect to ease of application, computational efficiency, and program capabilities. The personal computer used in this evaluation was a Hewlett Packard Vectra VL, Series 3 Pentium, 75 MHz, with a 420 MB hard disk and 16 MB of RAM. The personal computer is comparable to the high-end PCs available within the work environment that will be responsible for the analysis of the 1996 Medical Expenditure Panel Survey data, the next cycle of the NMES surveys. Because the NMES had a multistage sample design, the identification of the sampling unit levels was required for estimating standard errors. The first-stage sampling level variable (SSTRATA) identified the respective strata of the NMES design. The primary sampling units (e.g., counties) selected within strata were designated by the second-stage sampling level variable (SPSU). The sampling weight (INCALPER) was specified to allow for the computation of national estimates. The sampling weight adjusted for the disproportionate selection probabilities that characterized the NMES design, in addition to adjustments for survey nonresponse and poststratification adjustments.

The datasets that were used for the evaluation consisted of an SAS file containing 26 variables, which included 4 survey design variables (strata, PSU and person identifiers, sampling weight), 12 analytical variables, and 10 demographic measures. The data file that was used for the overall population estimates consisted of 34,459 individuals (dataset size: 5.620 MB), and the dataset that was restricted to the subset of individuals with medical expenses in 1987 consisted of 28,704 individuals (dataset size: 4.682 MB). The dataset included 101 unique strata and 202 PSUs specified for variance estimation, based on the NMES sample design. For all of the software procedures under consideration, the finite population correction options available in the respective programs were not used in this evaluation.

The version of Stata used for the evaluation was Intercooled Stata, Release 5.0 for Windows (StataCorp 1996; Eltinge and Scribney 1996). The procedure runs on an 80386 or better, and requires a math coprocessor, either on-chip (Pentium; 80486DX) or added (80486SX with 80487; 80386 with 80387). Stata 5.0 is available for Macintosh, and a Linux version and a DOS version are also available. The version of SUDAAN used for the evaluation was Release 7.0. (Shah, Barnwell, and Bieler 1995), and WesVarPC Version 2.02 (which is written in

Table 2. Sources of Payment for Health Care Services: Percent Source of Payment Distribution, United States, 1987

Population characteristic	Population in 1,000s	Out of pocket (s.e.)	Private insurance (s.e.)	Medicare (s.e.)	Medicaid (s.e.)	Other public programs (s.e.)	Other (s.e.)
Total	239,393	23.7% (.5)	36.2% (1.0)	18.1% (1.0)	8.3% (.6)	9.0% (1.3)	4.6% (.3)

Source 1987 NMES Household Survey, AHCPR.

C) within a Windows environment was used (Brick, Broene, James, and Severynse 1996). SUDAAN is also written in the C language, and the procedure runs on IBM-compatible PCs (MS-DOS and OS/2). A 386 or more powerful personal computer with a math coprocessor (4 MB of RAM) and at least 4 MB of hard disk space for file storage is recommended for WesVarPC applications, in addition to Windows 3.1 or Windows 95.

3.1 Programming Effort

In order to run the Stata program, it was necessary to first create an SAS transport file. Once the transport file was created using the SAS PROC COPY procedure, the SAS dataset was converted into a Stata datafile using the Stat/Transfer file transfer utility, Version 3.5 (Circle Systems 1995). In order to derive the necessary mean estimates and standard errors for the criterion variables specified in Table 1 for the overall population and the demographic subgroups, it was necessary to use 18 programming statements. In addition to producing the mean estimates and associated standard errors, the output included survey design effects (the ratio of the variance of the survey estimate based on the complex sample design divided by the variance derived under simple random sampling assumptions) and a 95% confidence interval. The procedure allowed for the specification of all of the desired criterion variables in a given statement request to produce overall survey means. Alternatively, separate program statements were required for each of the distinct demographic breakdowns that were required. When cross classifications of demographic measures were specified, separate statements were required to produce survey estimates for the marginals of the specified demographic measures. More specifically, the following program statements were required to obtain the required mean estimates and their standard errors for the estimates in Table 1 associated with the full NMES sample:

```
use C:\SUDTEST\BIN\STAT011.DTA
svyset pweight incalper
svyset strata sstrata
svyset psu spsu
svymean totalexp totalsp1 totalsp2
svymean totalexp totalsp1 totalsp2, by (age)
svymean totalexp totalsp1 totalsp2, by (smpsexr)
svymean totalexp totalsp1 totalsp2, by (race3)
svymean totalexp totalsp1 totalsp2, by (povstal)
svymean totalexp totalsp1 totalsp2, by (ratehlth)
svymean totalexp totalsp1 totalsp2, by (ssmsa)
svymean totalexp totalsp1 totalsp2, by (sregion)
svymean totalexp totalsp1 totalsp2, by (cendiv)
svymean totalexp totalsp1 totalsp2, by (povstal ratehlth)
svymean totalexp totalsp1 totalsp2, by (age race3)
svymean totalexp totalsp1 totalsp2, by (age smpsexr)
svymean totalexp totalsp1 totalsp2, by (race3 smpsexr)
svymean totalexp totalsp1 totalsp2, by (age race3 smpsexr)
```

Table 3. *Ease of Application—Number of Programming Statements*

Type of statistic	Stata	SUDAAN	WesVarPC
Means	*18*	*8*	*3*
Totals	*14*	a	a
Ratios	*14*	*9*	*3*
Totals: ratio numerator	*14*	b	b

[a] Included as output with mean estimates.
[b] Included as output with ratio estimates.

Only 14 commands are necessary (svytotal statements) to obtain the required estimates of population totals for the criterion variables specified in Table 1 and their standard errors for the overall population and the specified demographic classifications because the first four statements listed above do not need to be repeated. Similarly, when the dataset was restricted to the subset of individuals with a medical expense, a total of 32 programming statements was also necessary to produce the combined mean and total population estimates for this restricted population (the use and svyset statements were necessary to identify the appropriate dataset and survey design variables). When attention was directed to deriving the necessary ratio estimates and standard errors for the criterion variables specified in Table 2 for the overall population and the demographic subgroups, it was also necessary to use 14 programming statements (svyratio statements) and another 14 statements (svytotal statements) to yield the desired estimates for the corresponding population totals for the respective source of payments. Table 3 presents a summary of the programming statement requirements for the Stata program and the alternative software procedures.

The SUDAAN procedure was able to accept the SAS file with the NMES data directly as program input. The SUDAAN procedure required eight programming statements to produce the comparable mean estimates that were required for Table 1 for the full NMES sample. As in the Stata procedure, a single programming statement allowed for the specification of all of the desired criterion variables in a given statement request to produce survey means. In addition to the required mean estimates and their associated standard errors, the program allowed for the following additional requested output: population total estimates for specified criterion measures and their standard errors, sample size, and the survey design effects for the mean estimates. Only one program statement was required for each of the distinct demographic breakdowns that were required. Furthermore, the mean estimates and standard errors were also produced for cross classifications of the demographic measures that were specified and for their marginal values. The following program statements were required to obtain the required mean estimates and their standard errors for the estimates in Table 1 associated with the full NMES sample:

```
PROC DESCRIPT DATA=STAT1 FILETYPE=SAS;
NEST SSTRATA SPSU;
VAR TOTALEXP TOTALSP1 TOTALSP2;
SUBGROUP POVSTAL SREGION SSMSA RATEHLTH AGE
  RACE3 SMPSEXR CENDIV;
LEVELS 5 4 4 4 5 3 2 9;
WEIGHT INCALPER;
TABLES SREGION SSMSA POVSTAL*RATEHLTH CENDIV
  AGE*SMPSEXR*RACE3;
PRINT NSUM WSUM TOTAL MEAN SEMEAN SETOTAL
  DEFFMEAN/TOTALFMT=F16.0;
```

Similarly, when the dataset was restricted to the subset of individuals with a medical expense, a total of eight programming statements was also necessary. When attention was directed to deriving the necessary ratio estimates and standard errors for the criterion variables specified in Table 2 for the overall population and the demographic subgroups, it was necessary to use nine programming statements (PROC RATIO) to yield the desired ratio estimates in addition to the corresponding population total estimates and associated standard errors for the respective source of payments.

The WesVarPC program can also read the SAS files with the NMES data directly as program input. In order to run the WesVarPC procedure, however, it was necessary to create replicate weights necessary for deriving the estimated standard errors. For the purposes of this evaluation the replicate weights creation option in WesVarPC was used to create the replicate weights, with 100 replicates specified. The variance estimation procedure that was implemented was Balanced Repeated Replication (BRR), although the user has the option selecting one of two jackknife methods for variance estimation, or Fay's method (Brick et al. 1996). Three statements were necessary to facilitate the computation of the necessary replicate weights. The WesVarPC procedure also required three programming statements to produce the comparable mean estimates that were required for Table 1 for the full NMES sample. A single programming statement allowed for the specification of all of the desired criterion variables in a given statement request to produce a survey mean. In addition to the required mean estimates and their associated standard errors, the program allowed for the following additional requested output: population total estimates for specified criterion measures and their standard errors, sample size, the coefficient of variation, and the survey design effects for the specified survey estimates. Only one program statement was required for each of the distinct demographic breakdowns that were required. Furthermore, the mean estimates and standard errors were also produced for cross classifications of the demographic measures that were specified and for their marginal values.

```
ANALYSIS VARIABLES: TOTALEXP, TOTALSP1, TOTALSP2
COMPUTE STATISTIC; M_TOTALEXP=MEAN(TOTALEXP)
                   M_TOTALEXP=MEAN(TOTALSP1)
                   M_TOTALEXP=MEAN(TOTALSP2)
TABLE REQUESTS:    SREGION
                   SSMSA
                   POVSTAL*RATEHLTH
                   CENDIV
                   AGE*SMPSEXR*RACE3
```

Table 4. Approximate Execution Times for PC Software Packages to Produce Required Output

Type of statistic	*Stata*	*SUDAAN*	*WesVarPC*
Means (n = 34,459)	*18 minutes*	*<1 minute*	*15 minutes*
Total (n = 34,459)	*20 minutes*	a	a
Means (n = 28,704)	*13 minutes*	*<1 minute*	*12 minutes*
Totals (n = 28,704)	*14 minutes*	a	a
Ratios (n = 34,459)	*64 minutes*	*<1 minute*	*33 minutes*
Totals (n = 34,459)	*23 minutes*	b	b

[a] Included as output with mean estimates.
[b] Included as output with ratio estimates.

When the dataset was restricted to the subset of individuals with a medical expense, a total of three programming statements was also necessary. When attention was directed to deriving the necessary ratio estimates and standard errors for the criterion variables specified in Table 2 for the overall population and the demographic subgroups, it was necessary to use three programming statements to yield the desired ratio estimates in addition to the corresponding population total estimates and associated standard errors for the respective source of payments (Table 3).

3.2 Software Efficiency

The alternative software procedures were then compared in terms of computational efficiency as measured by execution time. Execution time was measured by the run time necessary to complete the program statements, from program submission to final result. Table 4 presents the approximate time to execute the respective software programs to produce the required estimates. Clearly, the execution time could be improved upon with a more powerful personal computer. The computer selected for the evaluation is comparable to the high-end PCs available within the work environment that will be responsible for the analysis of the 1996 Medical Expenditure Panel Survey data, the next cycle of the NMES surveys. It should serve to illustrate the relative efficiency of the respective procedures.

Overall, the SUDAAN procedure was consistently the most efficient procedure, requiring less than 1 minute of execution time to yield the required mean estimates, total estimates, and associated standard errors for each of three distinct program tasks:

1. derivation of mean and total estimates and their associated standard errors for the criterion variables specified in Table 1 for the full NMES sample ($n = 34,459$) and all specified subpopulations;
2. derivation of mean and total estimates and their associated standard errors for the criterion variables specified in Table 1 for the restricted NMES sample ($n = 28,704$) and all specified subpopulations;
3. derivation of ratio and total estimates and their associated standard errors for the criterion variables specified in Table 2 for the full NMES sample ($n = 34,459$) and all specified subpopulations;

Alternatively, the Stata procedure, which used the same Taylor Series linearization approximation method as SUDAAN to estimate the variance of the respective survey estimates, took the longest time to complete the required tasks, as noted by the computing time in excess of 1 hour to obtain the ratio estimates and their associated standard errors for the criterion variables specified in Table 1 for the full NMES sample ($n = 34,459$).

The WesVarPC procedure's use of a replication approach to variance estimation is largely responsible for the significantly greater length of execution time to produce the required output relative to the time required by the SUDAAN procedure. All three of the procedures were affected by the size of the dataset under consideration. A reduction in sample size from 34,459 to 28,704 resulted in a comparable relative reduction in the execution time necessary for the packages to produce the required survey estimates and associated standard errors.

3.3 Computation Accuracy

The estimates obtained from the alternative software procedures were compared, and resulted in equivalent mean, ratio, and population total estimates out to available decimal places, usually four places after the decimal point. The same equivalence was noted for the standard error estimates produced by the Stata and SUDAAN software procedures which both use the Taylor series approximation to compute variances. Furthermore, a comparison of the standard errors derived by using the WesVarPC procedures with those obtained by the Taylor series approximation yielded generally convergent results.

3.4 Program Capabilities

In addition to computing the required parameter estimates and their associated standard errors, the software procedures developed for the analysis of complex survey data in a PC environment possess a number of additional desired features in common, as well as a set of unique capabilities that distinguish them. All of the procedures offer the capacity to adjust the standard errors of the survey estimates by a finite population correction factor (Table 5). The Stata and the SUDAAN procedures include a capacity to adjust the variance of the survey estimates at the first stage of sample selection, with additional secondary and tertiary stage adjustment options available in SUDAAN. All of the procedures also allow for the computation of differences across subpopulation means, proportions, and ratios with the derivation of associated standard error estimates. In addition, all three procedures have the capacity to yield parameter estimates of model coefficients and their associated standard errors for both linear and logistic regression models. These regression procedures allow for hypothesis testing and for the derivation of p values associated with specific tests for model coefficients. All of the procedures have developed well-specified documentation to assist the user in running the specific program features. Each of the software packages has distinct strategies for handling missing data, and the users should consider this when choosing a package. The stata documentation is provided in the *Stata 5.0 user manual*, and the SUDAAN and WesVarPC documentation is provided in separate user manuals (StataCorp 1996; Shah et al., 1995; Brick et al., 1996).

With respect to the more hand-tailored features that are specific to the respective software procedures, Stata is unique among the set in offering an integrated statistical software package that builds a set of software procedures for complex survey data analysis into an existing statistical software package that features other statistical analysis, graphics, and data management capabilities. It is also the only statistical program available that will estimate the model coefficients and associated standard errors for probit models with adjustments for complex survey data. Another feature allows for the estimation of a linear combination of the coefficients obtained from linear regression, logistic regression, or probit models, as well as use of the Bonferroni procedure to test whether a group of coefficients is simultaneously equal to zero.

The SUDAAN procedure continues to build on its strengths as a family of statistical procedures designed for the analysis of complex survey data. With respect to descriptive statistics, SUDAAN also allows for the estimation of quantiles, relative risks, and linear contrasts and higher order trends, along with their standard errors adjusted for survey design complexities. The procedure also allows for the derivation of standardized estimates of means and percentages and their associated standard errors. In addition to its capacity to fit linear regression and logistic regression models and construct hypothesis tests concerning the model parameters, SUDAAN fits discrete and continuous proportional hazards regression models to failure time data and estimates hazard ratios and their 95% confidence intervals for model parameters. SUDAAN also offers a capacity to fit log-linear models to contingency tables, again adjusting for survey design complexities.

Through its MULTILOG procedure, SUDAAN also offers an extension of the modeling capabilities of SUDAAN to include categorical outcomes with more than two categories that may or may not have a natural ordering. The procedure implements the proportional odds model with cumulative logit link for ordinal responses and a generalized multinomial logit model for nominal responses. SUDAAN also groups the statistics that its procedures produce, and allows a capacity to write the statistics to ASCII, SAS, or SUDAAN output datasets for further analyses.

The WesVarPC procedure is the only widely utilized PC-based program available that uses the replication method for variance estimation. Replication techniques appropriately reflect the impact of a survey's nonresponse and poststratification adjustments in the variance estimates when used in tandem with replicate weights that also reflect these adjustments. The software procedure will also allow the user to create replicate weights for a multistage sample design, if they have not been developed. The system supports variable recodes, relabeling, and reformats, and will support the direct import of PC-SAS for DOS, SAS Transport, SPSS for Windows, DBF, or ASCII data files.

Another attraction of WesVarPC is its capacity to calculate variances for analysis variables involving complex transformations of the original survey variables and for regression parameters under consideration. This is a feature of replication techniques that allows users to define new variables and analytic functions of the variables, and to compute the variance of these statistics. It will also produce variance estimates having low relative bias for nonlinear statistics based on relatively small sample sizes. In addition, the WesVarPC output features allow the user to control the format of output files for direct input to a spreadsheet or other software for customizing the output or doing additional computations.

4. SUMMARY

All of the packages evaluated for this paper were straightforward to use, and included well-specified documentation to assist the user in both the more standard applications, in addition to implementing the more complex analytical features. The WesVarPC procedure consistently required the fewest

Table 5. *Comparison of Software Capabilities with Respect to Variance Estimation*

Feature	Stata	SUDAAN	WesVarPC
Means, proportions, totals, ratios; associated standard errors and design effects	*	*	*
Finite population correction	*	*	*
Detailed documentation	*	*	*
Quantiles		*	
Contrasts	*	*	*
Tests of independence		*	
Linear regression	*	*	*
Logistic regression	*	*	*
Logistic modeling for polychotomous outcome variables		*	
Proportional hazards model		*	
Log-linear modeling of contingency tables		*	
Probit modeling	*		
Output features	*	*	*

* indicates that the capability is available.

programming statements to derive the required survey estimates in each of the distinct comparisons under consideration, but required additional data preparation for the creation of replicate weights necessary for the derivation of variance estimates. Alternatively, even the more lengthy number of program statements necessary to obtain the same results through the Stata procedure provided no undue burden to implement, given Stata's windowed interactive interface. When attention was directed to computational efficiency, the SUDAAN procedure was consistently superior in generating the required estimates, requiring less than 1 minute of computer run time to produce each of the sets of estimates and standard errors that were required. Alternatively, the Stata procedure, which used the same Taylor Series approximation for variance estimates, took over an hour of computer run time to obtain the desired output.

All of the software packages under consideration were able to produce the specified NMES health care estimates, their related standard errors in an acceptable manner. The program output from each procedure facilitated the conduct of hypotheses tests that are characteristic of over 100 AHCPR analytical papers that have been prepared using the NMES data. In addition to the set of features they share in common, each package includes a set of unique capabilities that are hand-tailored to specific tasks. As noted, Stata builds a set of software procedures for complex survey data analysis into an existing statistical software package that features other statistical analysis, graphics, and data management capabilities. It is also the only statistical program available that will estimate the model coefficients and associated standard errors for probit models with adjustments for complex survey data. The single-user academic price for Strata 5.0 is $395, and the price for industry is $995 (Address: Stata Corporation, 702 University Drive East, College Station, TX 77840). The SUDAAN procedure continues to build on its strengths as a family of statistical procedures designed for the analysis of complex survey data. It has the most comprehensive set of analytical features available in a single package tailored to the analysis of complex survey data. The perpetual license fee for single personal computers is $1,195 for Release 7.0 for Windows 3.1, and $995 for Release 7.0 for DOS (Address: Statistical Software Center, Research Triangle Institute, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709-2194). In contrast, the WesVarPC procedure is the only widely utilized PC-based program available that uses the replication method for variance estimation. Implementing program specifically developed to take advantage of the Windows environment is particularly user friendly. It should also be noted that the WesVarPC procedure is the only package within this group that is made available to users at no charge by its developers. The software and documentation can be downloaded from the World Wide Web site at the following URL: (http://www/westat.com/wesvarpc/index.html) (address: Westat, 1650 Research Boulevard, Rockville, MD 20850-3129).

Based on the results of this evaluation, it is difficult to designate any of the procedures as summarily the "best" or "worst." The potential user should consider the attractions and limitations of the respective packages identified in this evaluation, in the context of their particular analytical requirements, in order to reach a decision regarding which package to acquire for the analysis of complex survey data.

[Received October 1996. Revised January 1997.]

REFERENCES

Barnwell, B. G., Bieler, G. S., and Shah, B. V. (1996), "SUDAAN Technical Report: The MULTILOG Procedure, Release 6.6," Research Triangle Institute, Research Triangle Park, NC 27709.

Brick, J. M., Broene, P., James, P., and Severynse, J. (1996), *A User's Guide to WesVarPC*, Rockville, MD: Westat, Inc.

Carlson, B. L., Johnson, A. L., and Cohen, S. B. (1993), "An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data," *Journal of Official Statistics*, 9, 795–814.

Circle Systems, Inc. (1995), *Stat/Transfer*, Seattle, WA.

Cohen, S. B. (1996), "The Redesign of the Medical Expenditure Panel Survey: A Component of the DHHS Survey Integration Plan," in *Proceedings of the COPAFS Seminar on Statistical Methodology in the Public Service*.

Cohen, S. B., Burt, V. L., and Jones, G. K. (1986), "Efficiencies for Variance Estimation for Complex Survey Data," *The American Statistician*, 40, 157–164.

Cohen, S. B., Xanthopoulos, J. A., and Jones, G. K. (1988), "An Evaluation of Statistical Software Procedures Appropriate for the Regression Analysis of Complex Survey Data," *Journal of Official Statistics*, 4, 17–34.

Cohen, S. B., DiGaetano, R., and Waksberg, J. (1991), "Sample Design of the 1987 National Medical Expenditure Survey—Household Component," AHCPR Publication 91-0037, National Medical Expenditure Survey Methods 3, Agency for Health Care Policy and Research, Rockville, MD: Public Health Service.

Eltinge, J. L., and Scribney, W. M. (1996), "Some Basic Concepts for Design-Based Analysis of Complex Survey Data," *Stata Technical Bulletin*, 31, 3–42.

Edwards, W., and Berlin, M. (1989), "Questionnaires and Data Collection Methods for the Household Survey and the Survey of American Indians and Alaska Natives," DHHS Publication (PHS) 89-3450, National Medical Expenditure Survey Methods 2, National Center for Health Services Research and Health Care Technology Assessment, Rockville, MD: Public Health Service.

Fuller, W. A., Kennedy, W., Schnell, D., Sullivan, G., and Park, H. J. (1989), *PC CARP*, Ames, IA: Statistical Laboratory, Iowa State University.

Newton, H. J. (1996), *Stata Technical Bulletin*, College Station, TX: Stata Corporation.

Shah, B. V., Barnwell, B. G., Bieler, G. S., Boyle, K. E., Folsom, R. E., LaVange, L., Wheeless, S. C., and Williams, R. (1996), *Technical Manual: Statistical Methods and Algorithms Used in SUDAAN Release 7.0*, Research Triangle Park, NC: Research Triangle Institute.

Shah, B. V., Barnwell, B. G., and Bieler, G. S. (1995), *SUDAAN User's Manual: Software for the Statistical Analysis of Correlated Data*, Research Triangle Park, NC: Research Triangle Institute.

StataCorp (1996), "Stata Statistical Software: Release 5.0," College Station, TX: Stata Corporation.

Wolter, K. X. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.