TECHNICAL NOTES

General Technical Background to the 1996 Health Status Survey

Introduction

The purpose of this section is to provide the reader with a general methodological overview of the project. Persons interested in obtaining additional or more detailed information may contact:

Bureau of Surveillance and Analysis
Office of Public Health Data
Utah Department of Health
P O Box 142101
Salt Lake City, UT 84114-2101
Phone: (801) 538-6108
E-mail: hlhda.phdata@state.ut.us

Sample Design

The 1996 Utah Health Status Survey represents the third such survey; previous surveys were conducted in 1986 and 1991. The statistical estimates in this report are based on *1996 Utah Health Status Survey* data.

The sample was a **complex survey sample** designed to be representative of all Utahns. It is best described as a weighted probability sample of approximately 6,300 households disproportionately stratified by twelve local health districts that cover the entire state. Five hundred household interviews were conducted in each health district, except Salt Lake City/County Health District, in which eight hundred household interviews were conducted in order to increase the precision of statewide estimates.

A **single stage, non-clustered, equal probability of selection telephone calling design** was used to generate telephone numbers, more specifically referred to as the *Casady-Lepkowski* (1993) calling design. This method begins by building a *base sampling frame* consisting of all possible telephone numbers from all working prefixes in Utah. Telephone numbers are arranged sequentially into groups of 100 by selecting all telephone numbers within an area code and prefix, plus the first and second digits of the suffix (e.g., 801-538-10XX represents a group that includes all 100 phone numbers between 801-538-1000 and 801-538-1099). Each group of 100 telephone numbers is classified as either high density (at least one residential listing) or low density (no listed residential phone numbers in the group). All low density groups are removed, and high density groups are retained. Telephone numbers are randomly selected from the high-density list. This sampling design ensures that both listed and unlisted phone numbers are included in the sample.

The survey interview was conducted with **one randomly-selected adult** (age 18 or older) in each household. To select this person, Gallup interviewers collected household membership information from the household contact person (the person who answered the telephone). One household member was then selected at random from the list of all household members age 18 or over. Survey questions were then asked about either, 1) all household members, 2) the survey respondent only, 3) a randomly selected adult or child household member (selected using the same method as was used to select the respondent), or 4) the household as a whole. Thus, the survey sample varies, depending on the within-household reference sample that was used for each set of survey questions. Each within-household reference sample has known probabilities of selection and can be generalized to the Utah population.

Survey Data Collection

The Utah Department of Health contracted with The Gallup Organization to collect the survey data. Gallup incorporated the telephone survey instrument into a **computer-assisted random digit dialing software program,** called SURVENT. Interviews were conducted by trained interviewers in a supervised environment across six sites. Interviews were conducted in Spanish when appropriate.

Computer-assisted telephone interviewing was chosen as the method of data collection for several reasons. First, it yields higher response rates, thus resulting in a more representative sample and reducing the amount of bias inherent in mail survey response rates. Second, it helps reduce non-sampling error by standardizing the data collection process. Data-entry errors are reduced because interviewers are not allowed to enter non-valid codes. It was also efficient because it allowed interviewers to enter responses directly into the database.

The survey questionnaire was divided into ***core*** and ***supplemental modules***. Core questions were asked of all households in the sample. Table 1 describes the types of "core" questions that were asked, and about whom they were asked. Notice that *not all questions were asked with regard to everyone in the household*.

Table 1.
CORE MODULE QUESTIONS

Question Topic	Within-Household Reference Sample
Demographic characteristics	All household members
Presence of chronic medical condition	All household members
Health insurance status	All household members
Injury incidence/safety issues	All household members
Lifestyle (smoking, drinking, exercise)	All household members
Subjective mental/physical health (SF12)	Respondent only (randomly-selected adult)
Health screening exam usage	Respondent only (randomly-selected adult)
Access to care/primary provider	Randomly-selected household member of any age
Household-level demographic characteristics	The household as a whole

In addition to the core survey questions (above), one of six different *supplemental modules* was administered to primarily non-overlapping randomly-assigned subsets of (approximately 1,000) households. Table 2 shows the types of questions asked in the supplemental module questions, and about whom they were asked.

Table 2.
SUPPLEMENTAL MODULE QUESTIONS

Type of Question	Within-Household Reference Sample
Limitations of activities	All household members
Migration	Respondent only (randomly-selected adult)
Health Plan Consumer Satisfaction	Respondent only (randomly-selected adult)
Fertility	Respondent or spouse only
Health Care Utilization	Randomly-selected household member of any age
Interpersonal violence	The household as a whole

*Note: All supplemental module questions were asked only of a subset of households.

While both core and supplemental modules yielded sufficient sample sizes to construct state-level estimates for the Utah population, the information collected from supplemental modules was not intended for use in district-level analyses.

Cooperation rate

The interview process took place over a three month period (from June to August, 1996), and resulted in a cooperation rate of 66.3%. If necessary, up to nine telephone attempts were made to contact a selected household. After a randomly-selected survey respondent was identified, up to nine attempts were made to conduct the interview with that person.

Weighting and Estimation Methods

Post-survey weighting adjustments were made so that the Health Status Survey findings could be more accurately generalized to Utah's population. Two types of post-survey weighting adjustments were made, one that adjusted for random sampling variation, and one that adjusted for disproportionate sampling (such as the over-sampling of smaller local health districts across the state). Although the two types of adjustment are distinct conceptually, they were accomplished in a single step.

The post-survey weighting adjustments weighted the sample to be proportionately consistent with the age, sex, geographic, and Hispanic status distribution of the 1996 Utah population. Utah population estimates by sex, single year of age, and county of residence were provided by the Utah Governor's Office of Planning and Budget (GOPB) (the estimates used were those compiled in 1994). Estimates of Utah's Hispanic population for 1996 were derived by calculating the average annual rate of increase of Hispanic persons for each health district using data from 1990 to 1994 Bureau of the Census reports, and then projecting those increases to 1996 GOPB local health district population counts. Total state estimates for Hispanic persons were calculated by summing across local health districts.

The post-survey weighting variables adjusted for the following factors:

1. The number of **phones** in the household.
2. The total **number of persons in the household** to which the data will be generalized (1 for questions that were asked about every household member, the number of adults in the household for questions that were asked only of the respondent, the number of persons in the household for questions that were asked of a randomly-selected household member).
3. The proportion of **Hispanic persons** in each local health district.
4. The **age and sex** distribution of each local health district.
5. The probabilities of selection for each **local health district**.

Population count estimates. Once a percentage was calculated for a variable of interest (e.g., the percentage uninsured) using appropriately weighted survey data, a population count (N) to which the percentage applied was estimated. In some cases analyses referenced certain age or sex groups, Hispanic persons or combinations of Utah counties. These total population group counts were readily available from the sources described earlier. However, for other groups where population counts were largely unavailable (e.g., analyses that examined the distribution of adult males by marital status), the population counts were estimated. This was achieved by multiplying the appropriate 1996 population total for that group (from 1996 GOPB estimates) by a proportion obtained from a frequency distribution or cross tabulation analysis of survey data. For instance, to calculate a population count for adult males who were married, the population of adult males from GOPB was multiplied by percentage of married adult males in the 1996 Utah Health Status Survey sample. Thus, any population count estimates not derived directly from existing age, sex, Hispanic status or county population estimates were derived from 1996 Health Status Survey data, and must be considered estimates.

Missing Values. Another consideration that affected the presentation of the population estimates in table format was the inclusion or exclusion of missing values ("don't know" and "refused to answer"). Population percentage estimates were calculated after removing the "don't know" and "refused to answer" responses from the denominator. This, in effect, assumed that persons who gave these answers were distributed identically on the variable of interest to those who gave a valid answer to that variable. For instance, that among those who did not know whether they were insured, we assumed that 90.47% of them were insured and 9.53% were not insured -- percentages identical to those found among the sample members who answered the question with a valid response.

Removing the missing cases from an analysis is rather simple and straightforward for analyses of a single variable. However, when one variable is cross-tabulated by another variable, all missing cases from both variables must be removed from the analysis. Removing the missing cases in itself is not a problem. However, a problem is encountered when a population estimate for a given variable, such as the percentage of all Utahns that have health insurance, differs slightly from an analysis of "all Utahns" versus an analysis of "all Utahns by age group." This is because the missing cases on the age variable have been removed from one analysis and not from another. Since the percentage of all Utahns that have health insurance was calculated on slightly different samples, the resulting percentage estimates are slightly different. This problem was resolved by reporting the best population estimate available for any given population subgroup. For instance, in the table of insurance rates for all Utahns by age, the population estimate from an analysis that includes Utahns of all ages, regardless of whether they reported missing values on the age variable has been substituted for the original total row in that table. The only drawback to this strategy is that the population count figures for Utahns with and without health insurance in tables like the "Utahns by Age Group" table do

not sum to the same number derived from the analysis of all Utahns regardless of whether they had missing values on the age variable. As a result, the tables appear as though they do not "add up."

Limitations and Other Special Considerations

Estimates developed from the sample may differ from the results of a complete census of all households in Utah due to two types of error, sampling and non-sampling error. Each type of error is present in estimates based on a survey sample. Good survey design and data collection techniques serve to minimize both sources of error.

Sampling error refers to random variation that occurs because only a subset of the entire population is sampled and used to estimate the finding in the entire population. It is often mis-termed "margin of error" in popular use. Sampling error is expressed as a *confidence interval.* The 95% confidence interval (calculated as 1.96 times the standard error of a statistic) indicates the range of values within which the statistic would fall 95% of the time if the researcher were to calculate the statistic (e.g., a percentage) from an infinite number of samples of size=n drawn from the same base population. It is typically expressed as the "plus or minus" term, as in the following example:

> "The percentage of those polled who said they would vote for Bill Clinton was 52%, plus or minus 2%."

Because local health districts were disproportionately stratified and then weighted to reflect the Utah population, the sample was considered a complex survey sample design. Estimating the sampling error for a complex survey design requires special statistical techniques, derived from the standard error for each estimate. SUDAAN software (Research Triangle Institute) was chosen to estimate the standard errors of the survey estimates because it employs a statistical routine (Taylor-series expansion) that accounts for the complex survey design.

Figures in this report include bars showing this estimated confidence interval around the parameter estimate. In cases where the confidence interval was greater in magnitude than the estimate, the estimate was not given. Estimates were not computed where the sample denominators were less that n=50. Readers should note that we have always presented the confidence interval as though it were symmetric, that is, of equal value both above and below (plus and minus) the estimate. It is often the case, however, that a confidence interval will be nonsymmetric. This occurs when the distribution is positively or negatively skewed, such as when a percentage is close to 0% or 100%. However, because the software program we use provides only symmetric confidence intervals, we are unable to provide the asymmetric estimates.

Non-sampling error also exists in survey estimates. Sources of non-sampling error include idiosyncratic interpretation of survey questions by respondents, variations in interviewer technique, household non-response to questions, coding errors, and so forth. No specific efforts were made to quantify the magnitude of non-sampling error.

Comparability with other surveys is an issue with all surveys. Differences in survey design, survey questions, estimation procedures, the socio-demographic and economic context, and changes in the structure and financing of the health care delivery system may all affect comparison between the 1996 Utah Health Status Survey and other surveys, including those conducted by the U.S. Bureau of the Census, the Behavioral Risk Factor Surveillance System surveys, and previous Utah Department of Health, Health Status Surveys.

Telephone surveys exclude certain population segments from the sampling frame, including persons in group living quarters (e.g., military barracks, nursing homes) and households without telephones. At the time of the 1990 Decennial Census, only four percent of Utah households were without telephone service. Typically, telephone surveys are biased because telephone households under-represent lower income and certain minority populations. In addition, studies have shown that non-telephone households tend to have lower rates of health care utilization (especially dental care), poorer health habits and health status, and lower rates of health insurance coverage (Thornberry and Massey, 1988).

Despite these overall disparities between telephone and non-telephone households, new survey research (Keeter, 1995) suggests that a similarity exists between data from non-telephone households and telephone households that experienced an interruption in service over the past 12 months. This similarity exists because many, if not most, households currently without telephones did have service in the recent past, and will have service again in the future. Therefore, certain households with telephones (those that had a recent interruption in service) are representative of "nonphone" households, allowing health status survey estimates that have been corrected for telephone noncoverage bias to be produced where indicated.