# Technical Notes

## General Technical Background to the 2003 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:

Office of Public Health Assessment
Center for Health Data
Utah Department of Health
P O Box 142101
Salt Lake City, UT 84114-2101
Phone: (801) 538-6108
E-mail: phdata@utah.gov

Sample Design

The 2003 Utah Health Status Survey represents the fifth such survey: previous surveys were conducted in 1986, 1991, 1996, and 2001. The statistical estimates in this report are based on *2003 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 3,175 households disproportionately stratified by twelve local health districts that cover the entire state. The sample was stratified so that the survey estimates could be provided for each local health district.

| | | Unweighted Counts | |
|---|---|---|---|
| | Health District/Small Area | Households | Persons |
| 1 | Bear River Health District | 201 | 675 |
| 2 | Central Utah Health District | 210 | 664 |
| 3 | Davis County Health District | 196 | 658 |
| 4 | Salt Lake Valley Health District | 704 | 2,163 |
| 5 | Southeastern Utah Health District | 213 | 621 |
| 6 | Southwest Utah Health District | 204 | 618 |
| 7 | Summit County Health District | 195 | 550 |
| 8 | Tooele County Health District | 233 | 755 |
| 9 | TriCounty Health District | 205 | 590 |
| 10 | Utah County Health District | 390 | 1,379 |
| 11 | Wasatch County Health District | 214 | 691 |
| 12 | Weber-Morgan Health District | 210 | 594 |
| | State Total | 3,175 | 9,958 |

A **single stage, non-clustered, equal probability of selection telephone calling design**, more specifically referred to as the Casady-Lepkowski (1993) calling design, was used to generate telephone numbers in each local health district. 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 Utah Department of Health Survey Center collected the survey data. The survey interview was conducted with **one randomly selected adult** (age 18 or older) in each household. To select this person, interviewers collected household membership information from the household contact person (the person who answered the phone). The adult household member who had celebrated the most recent birthday was then selected from the list of all household members age 18 or over. Survey questions were then asked of the respondent about either, 1) all household members, 2) the survey respondent only, 3) a randomly selected adult or child household member (used only in the injuries section), or 4) the household as a whole. Data were collected on all household members through the respondent. Thus, the survey sample varies, depending on the within-household sample that was used for each set of survey questions. Each within-household sample has known probabilities of selection and has been weighted appropriately so it can be generalized to the Utah population.

*Questionnaire Construction*

The 2003 Utah Health Status Survey was based on the 2001 and 1996 Utah Health Status Survey questionnaires. For the 2003 questionnaire, some changes were made based on input from the Health Surveys Advisory Committee and the Health Status Survey staff. These changes were made in order to obtain more detailed information and to allow for comparison with large, federal surveys, such as the Current Population Survey (CPS). The entire survey questionnaire may be found on-line at http://health.utah.gov/ibis-ph/opha_pubs.html.

*Survey Data Collection*

The Utah Department of Health Survey Center integrated the survey questionnaire into a **computer-assisted telephone interviewing (CATI)** software program. Interviews were conducted by trained interviewers in a supervised and monitored environment at the Utah Department of Health Survey Center.

**Computer assisted telephone interviewing** was chosen as the method of data collection for several reasons. First, it yields high 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.

*Response Rate*

The interview process took place over a ten-month period (from March to December, 2003), and resulted in a response rate of 65.9%. If necessary, up to fifteen telephone attempts were made to contact a selected household.

# Technical Notes

*Weighting 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 the smaller local health districts across the state). Although the two types of adjustments are distinct conceptually, they are accomplished in a series of steps that does not distinguish between the two types.

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

1. The number of **phone lines** in the household.
2. The total **number of adults in the household** (for questions that were asked only of the respondent, but were meant to be generalized to all adults in the household).
3. The proportion of **Hispanic persons** in each local health district.
4. The population **age and sex** distribution of each local health district.
5. The probabilities of selection for each **local health district**.

*Calculation of Survey Estimates*

**Population count estimates**. Once a percentage was calculated for a variable of interest (e.g., the percentage uninsured) using appropriately weighted survey data, it was applied to a population count to derive the estimate for the number of Utahns affected. In some cases analyses referenced certain age or sex groups, Hispanic persons or combinations of Utah counties. The population count estimates for these groups were readily available from the Utah Governor's Office of Planning and Budget. However, for other groups where population counts were largely unavailable (e.g., analyses that examined the distribution of adult males by marital status), survey data were used to estimate the population counts. This was achieved by multiplying the appropriate 2003 population total for that group (from 2003 GOPB estimates) by a proportion obtained from a frequency distribution or cross tabulation analysis of Utah Health Status Survey data. For instance, to calculate a population count for adult males who were married, the population of adult males from GOPB estimates was multiplied by percentage of married adult males in the 2003 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 2003 Health Status Survey data, and were rounded to the nearest 100 persons.

**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, assumes that persons who gave those 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 90.9% of them were insured and 9.1% were not insured—percentages identical to those found among the sample members who answered the question with a valid response.

Readers may have noticed that the numbers in the last two columns of the reference tables do not always sum to the total as they should. This was unavoidable for two reasons:

1) If there were missing values on the demographic grouping variable, the sum of the parts is derived from a slightly different sample than the estimate for the overall number.
2) The post-survey weighting adjustments cause certain irregularities in the tables.

*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, or parameter, in the entire population. It is often termed "margin of error" in popular use. Sampling error has been expressed in this report 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 the same size 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 George W. Bush was 47%, plus or minus 2%".

Because the sample was clustered within households, and because local health districts were disproportionately stratified and then weighted to reflect the Utah population, the sample is considered a complex survey sample design. Estimating the sampling error for a complex survey design requires special statistical techniques. SAS software, using "proc surveymeans," was used 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 error 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. Additionally, estimates were not computed where the sample denominators were less than 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 have not provided 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. Non-sampling error was minimized by good questionnaire design, use of standardization in interviewer behavior and frequent, on-site, interviewer monitoring and supervision.

**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 2003 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, such as persons in group living quarters (e.g., military barracks, nursing homes) and households without telephones. At the time of the 2000 Decennial Census, only two 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, the Utah Health Status Survey estimates may be considered adequately representative of all Utah households. Certain 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 "non-phone" households, allowing health status survey estimates to be corrected for telephone non-coverage bias. This correction has typically not been made, and will be clearly indicated when it is used.