Technical Notes

General Technical Background to the 2001 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 2001 Utah Health Status Survey represents the fourth such survey: previous surveys were conducted in 1986, 1991, and 2001. The statistical estimates in this report are based on *2001 Utah Health Status Survey* data.

The sample was a **complex survey sample** designed to be representative of all non-institutionalized Utahns in households with telephones. It is best described as a weighted probability sample of 7,520 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

	Unweighted Counts	
Health District / Small Area	Households	Persons
1 Bear River Health District	619	1,985
2 Central Health District	476	1,537
3 Davis County Health District	470	1,565
4 Salt Lake Valley Health District	1,615	5,110
5 Southeastern Health District	484	1,403
6 Southwest Health District	501	1,576
7 Summit Health District	510	1,513
8 Tooele Health District	611	2,030
9 Tri-County Health District	587	1,862
10 Utah County Health District	763	2,691
11 Wasatch Health District	453	1,518
12 Weber/Morgan Health District	431	1,298
State Total	7,520	24,088

for each local health district.

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 contracted with PEGUS Research Inc. to collect the survey data. The survey interview was conducted with **one randomly selected adult** (age 18 or older) in each household. To select this person, PEGUS 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 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. 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 2001 Utah Health Status Survey was based on the 1996 Utah Health Status Survey questionnaire. For the 2001 questionnaire, some changes were made based on input from the Health Surveys Advisory Committee and the Health Status Survey staff. These changes included enhancing the sections on health insurance coverage and access to health care. 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://ibis.health.utah.gov/ophapubs.html.

Survey Data Collection

PEGUS Research, Inc. incorporated the telephone survey instrument into a **computer-assisted telephone interviewing (CATI)** software program. Interviews were conducted by trained interviewers in a supervised and monitored environment at one location in the Salt Lake Valley. One hundred and eighty-five interviews (2.5%) were conducted in Spanish.

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 seven-month period (from May to November, 2001), and resulted in a response rate of 40.8%. If necessary, up to fifteen telephone attempts were

made to contact a selected household.

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, 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. The population count estimates for these groups were readily available from the 2000 Census. 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 2000 population total for that group (from 2000 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 2001 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 2001 Health Status Survey data.

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 91.3% of them were insured and 8.7% 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 the 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 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 2001 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. Typically, telephone surveys are biased because telephone households underrepresent 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. The 2000 U.S. Census indicated that only 2% of Utah households were without telephone service in April of 2000. Furthermore, 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.

Analysis of the SF-12 Scale

The purpose of this section is to provide a more thorough presentation of the methodology that was used to compute the SF-12 physical and mental composite scales and difference scores used in this report. Readers who are interested in using the SF-12 items should contact QualityMetric Incorporated. For further information about QualityMetric Incorporated go to http://www.qualitymetric.com.

This section is intended to provide only additional information that pertains specifically to the Utah administration of the SF-12 in the context of the 2001 Utah Health Status Survey. General information on the administration of the 2001 Utah Health Status Survey may be found in the section entitled General Technical Background to the 2001 Health Status Survey.

Brief Background of the SF-12

The SF-12 is a self-reported measure of a person's perceived health on a number of dimensions (e.g. general health status, pain, depression, etc.). It was designed to measure patient outcomes in medical practice and clinical research for a variety of purposes and has been used to measure health outcomes among groups with various physical and mental disorders, as well as compare the health outcomes of different sociodemographic groups (e.g. sex, age, education, poverty status, marital status). The Medical Outcomes Study group developed the SF-12 with the following objectives in mind 1) to serve as a measure of overall health status that took the patient's perspective into account, 2) to meet the need for a standardized health status measure-

ment tool that was comprehensive, psychometrically sound, and brief (Ware & Sherbourne, 1992).

The SF-12 Version 2 is the most recent in a series of health status measures developed by the Medical Outcomes Study Group. Early on there were 18 and 20 item measures. More recently, a 36 item short form health status scale (SF-36) has replaced the earlier versions. The SF-36 can be scored to yield two overall measures: physical health and mental health summary measures. Each measure is composed of eight subscales, representing eight different dimensions of physical and mental health:

Physical functioning,
Role functioning (physical),
Bodily pain,
General health,
Vitality,
Social functioning,
Role functioning (emotional), and
Mental health

All eight subscales (36 items) are used to form both the physical and mental health summary measures. The first four dimensions are weighted more heavily in the construction of the physical health summary score (PCS), while the second four dimensions are weighted more heavily in the construction of the mental health summary score (MCS). The SF-36 can discriminate relatively well between persons with minor medical conditions, serious physical conditions, psychiatric conditions, and those with both serious physical and psychiatric conditions (Ware et al., 2000).

The SF-12 is not intended to replace the SF-36. Rather, a subset of 12 questions was selected from the SF-36, because 36 items were too many to include on most questionnaires (the 2001 Utah Health Status Survey being no exception). The 12 item subset explains over 90% of the statistical variance in the original 36 item physical and mental health summary scale measures. It can be scored so that it reproduces the average scores for the summary measures with a high degree of comparability, and it can be printed on one to two pages of a self-administered questionnaire or administered by an interviewer in less than two minutes on average (Ware et al., 2002).

Differences in SF-12 Versions One and Two

Version two of the SF-12, which was used in this study, differs in a few ways from version one. The second version of the SF-12 uses the same basic 12 questions used in the first version, however, changes were made to the layout of questions and response categories to improve readability and completion rates. Version two has greater comparability with translations and cultural adaptations that are widely used in the U.S. and other countries. Five-level response categories replaced previous response options for the Role Emotional and Role Physical subdomains. These changes were made to extend the range measured and increase score precision without increasing respondent burden. More specifically, changes in response options resulted in:

- a four-fold increase in the number of levels defined;

- more than five-fold increase in the range measured;
- substantially smaller standard deviations; and,
- a substantial reduction in the percentages of respondents who score at the ceiling and floor.

Five-level response categories were also used to replace the six level response categories used in the Mental Health and Vitality subdomains. The decision to eliminate one of the six response choices ('a good bit of the time') was based on research using the Thurston Method of Equal-Appearing Intervals (Thurston, 1929). Eliminating one of the response categories simplified the format of the question with little or no loss in information. In spite of changes made to the second version of the SF-12, the two versions are directly comparable to each other because the same methods were used to score and weight the measures to create the Physical and Mental Composite Scales (PCS & MCS) (Ware et al, 2002).

Data Collection

The Utah Health Status Survey interview began with a set of questions about the general characteristics of each household member (e.g. age, height, weight, race). One SF-12 item, (In general, would you say your/[name's] health is poor, fair, good, very good or excellent?) was asked for all household members. The remaining SF-12 questions were asked immediately after the general demographic questions to avoid the context effects that other material in the survey (e.g., questions on chronic conditions and doctor visits) might have upon responses to the SF-12 questions. With the exception of the general health question, the remaining SF-12 questions were asked only of the survey respondents. The respondent was not asked to provide information on other persons in the household because it was believed that he or she could not accurately provide proxy data for other household members for the SF-12 items. As a result, the SF-12 results in this report were derived from the responses of the 7,520 randomly selected adult respondents, and are representative of persons age 18 and over in the State of Utah. They do not, however, represent the health status of those under the age of 18.

Data Analysis

Initial Scoring. The SF-12 items were scored according to the procedure in Ware et al (2002). All items were coded so that high scores represented higher degrees of health. For example, the question about general health (In general, would you say your health is excellent, very good, good, fair or poor?) was scored so that 1 indicated poor health and 5 indicated excellent health. Several items had to be reverse coded from the way they were originally asked to obtain this order for all SF-12 items. (A detailed description of this process may be obtained from the Center for Health Data Office of Public Health Assessment).

The weighting algorithm was designed so that the SF-12 scores were consistent with both the SF-36 and SF-12 Version 1 scores, that is, each had a national mean of 50 and standard deviation of 10. Scores higher than the mean indicated that a person has better health status than average, while scores lower than the mean indicated poorer health status than average.

In Utah, as in national samples of the general population, the distribution was negatively skewed, with a range of approximately 10 to 70. Given this distribution of scores, persons with poor health outcomes tended to score much lower than the mean, as many as 40

points lower, but persons with excellent health outcomes tended to score only as many as 20 points above the mean.

Age-Specific Difference Scores. The physical and mental health summary measures differ by age group, with older persons experiencing worse physical health, but better mental health than those younger. Because of this pattern of responses, it is recommended that a person's score be interpreted in the context of his or her own age group. In order to compare across various population groups while controlling for the effects of age, we created a single score that would take into account age differences in responses (see Table 1).

Age-specific difference scores were calculated as a response to both the need for a single score that controlled for the effects of age, and the need for a scale that is more intuitive. According to psychometric scaling theory, a scale is a much more powerful measurement tool than a single item. Single items are prone to error, such as differences in interpretation by respondents. A scale is also advantageous because it can measure more of the richness of a phenomenon, such as measuring all eight dimensions of health status, ensuring that the full range of experiences is represented in the data. However, scales also have the disadvantage of often being less intuitive than a single item. It is difficult to know what a person's SF-12 score means. For example, just knowing that a given person has a PCS score of 42.5 does not tell a lot to most users of SF-12 data.

The age-specific difference score is perhaps the most intuitive way to understand a person's score. By looking at a difference score, it is immediately clear whether a person is healthier or less healthy than other persons in their comparison group. The age-specific difference score is the difference between a person's score and his or her age-specific reference group mean. Thus, if a person has a difference score of -5.5, it indicates that they scored 5.5 points lower than other persons their age -indicating somewhat poorer health. Additionally, difference scores can be compared across age groups, that is, a score of -5.5 means the same thing, regardless of the person's age. Another advantage of difference scores is the ability that the scores provide in comparing across sociodemographic groups. For example, this report used the difference scores extensively to measure the relative effect that chronic conditions such as diabetes or asthma have for persons across groups based on factors like age, gender, educational level, poverty status, and household income.

Developing Cut-Points for Above and Below Average. After computing the age-specific difference scores, the SF-12 scales were more intuitive than they had originally been. Positive scores indicated good health, while negative scores indicated poor health. However, there was still a question of how low a person's score had to be in order for him or her to be considered in poor health. Difference scores indicate the direction and magnitude of the score, but do not indicate at which point a negative score should cause concern.

The standard error of measurement is used to assign cut-off points for individual scores. The SEM is a psychometric property of the scale that indicates the extent to which an individual's score should be expected to vary over a large number of randomly parallel tests (given that his or her health has not changed) (Kosinksy, 1997; Ware, Bayliss, Robers, Kosinski & Tarlov, 1996; Nunnally, 1978). It is computed as follows:

SEM=standard deviation * (sqrt (1 - reliability coefficient))

The unweighted sample data were used to compute the reliability coefficient (also known as Cronbach's alpha). Weighting the survey sample was seen as unnecessary for this step because the reliability coefficient is a property of the scale that is based on the intercorrelation of items-we were not producing an estimate of a population parameter that would be generalized to the state population. In practice, weighting the data made very little difference in the value of the reliability coefficient. The same reliability coefficient was used to compute the SEM for both physical and mental health summary measures because all 12 items are used in the computation of both scales.

Weighted sample data were used to calculate the standard deviation for the two scales. Neither SUDAAN nor SAS Proc Survey Means were used to calculate the standard deviation. Standard deviations for the two scales were 9.82 and 8.79 for the PCS and MCS, respectively.

The standard errors of measurement for the physical and mental health scales were multiplied by 1.96 to derive the 95% confidence interval, the theoretical range of values within which an individual's score would vary over 95% of a large number of repeated observations with parallel forms of the same test. Conceptually, this confidence interval includes the mean scale score. If the confidence interval for a given person includes the average score, then he or she should be considered 'no different from average.' In practice, however, the confidence interval can be applied to the mean scale score to define a range, within which an individual score would be considered average. The 95% confidence intervals for the physical and mental health summary scores were 6.97 and 6.24, respectively.