Technical Appendix

Designation of Small Area Boundaries

ZIP codes and counties were used individually or combined to create 61 geographic areas. ZIP code areas were primarily used to define small areas in the current study because they are the smallest commonly-used geographic units that are also identified in most health data sources. ZIP code areas are discrete geographic areas used by the U.S. Postal Service in mail delivery that often roughly follow political boundaries. In some sparsely populated areas, counties were used as the geographic unit of interest.

Population size criteria for designing the small areas in this study were determined based on health event incidence rates. Smaller areas may be more meaningful to communities, but rates based on small numerators are unstable (Buescher, 1997) and confidence intervals for such rates are large, rendering the comparisons uninterpretable for most practical purposes. Using such small areas with small numbers of events may also pose privacy problems for more sensitive events, such as suicide or AIDS. The population size criteria were determined by examining the three- and five-year incidences of selected events, such as infant mortality and lung cancer, for which small area estimates were desired. A numerator of 20 or greater produces relatively stable estimates, and also simplifies computation of confidence intervals from a Poisson distribution (Ahlbom, 1993). It was determined that areas with 40,000 to 60,000 persons would produce incidence counts of 20 or more for a wide range of health events. Increasing the population sizes sufficiently to produce reliable estimates for rare events (e.g., homicide or AIDS) would increase area size beyond that which would allow meaningful community level analyses. Where possible, areas with 40,000 to 60,000 persons were established, but areas with population sizes of approximately 20,000 were created when low population density, community identity, or others factors suggested that it was appropriate.

Areas were geographically constrained so that their boundaries would not cross local health district boundaries. Whenever possible the following conditions were met: local health districts were divided into multiple small areas, only contiguous ZIP codes were combined, sub-county small areas were not combined with ZIP code areas in neighboring counties, areas conformed to established political boundaries of cities and towns, and a ZIP code area was not combined with another area with an extremely different estimated median income. For some areas (primarily the urban counties that were subdivided into many small areas), a draft of the small area design was submitted to local representatives. The local representatives (10 of the 12 Utah local health officers, and 26 city officials selected from the directory of the Utah League of Cities and Towns) were provided a map of their locality showing the proposed small area boundaries and asked to consider whether the combined ZIP code areas were similar in terms of lifestyle and demographic characteristics. Several changes were made based on their recommendations.

Population Estimates

Data on population size, median age, and median income were purchased for current Utah ZIP codes from a commercial vendor, CACI Marketing Systems. CACI constructed population estimates at the ZIP code level by using the most recent decennial census data and additional information, such as sub-county estimates of change from the U.S. Census Bureau, special censuses, local sources of information about change, and changes in residential delivery statistics from the U.S. Postal Service. Estimates included 1997 population totals and population by sex and age group for each ZIP code. The age-specific population estimates were used in age-adjusting the data. The CACI file also included estimates for the average annual rate of population change for each ZIP code area, which were used to derive the 1994 through 1996 population estimates required for these analyses.

Selection of Measures

Sixteen measures were selected for presentation in this report. The 1997 estimates of two demographic variables, median age and per capita income, were purchased from CACI along with the population estimates. Nine variables, births to adolescents, low birth weight, prenatal care, infant mortality, deaths from all causes, motor vehicle crash deaths, suicide, lung cancer deaths, and cardiovascular disease deaths, are Health Status Indicators that were developed as part of the Healthy People 2000 process under the leadership of the Centers for Disease Control and Prevention. These nine health status indicators were selected because the events occurred with sufficient frequency to be meaningful at the small area level. Birth rate and general fertility rate were selected for use from the birth certificate data set. Finally, three variables, overall health status, cigarette smoking, and health insurance coverage, were selected from the 1996 Utah Health Status Survey data set because they were good indicators of overall health status, and the data were available for over 20,000 household members included in the survey.

Calculation of Rates

Typically, the number of events (e.g., number of deaths) in a given area has little meaning unless the size of the population is known. A rate is a fraction in which the numerator is the number of events, and the denominator is the number of people in the population at risk over the same period of time. For example, there were 13 infant deaths in the Brigham City area from 1992 through 1996, and 1,708 births: 13 / 1,708 = .0076 infant deaths for every live birth in Brigham City from 1992 through 1996. Small fractions are generally communicated as multiples of 100 (i.e., a percentage) 1,000, or 100,000. In the example above, we could say that there were 7.6 infant deaths for every 1,000 live births in Brigham City over the time period.

For some measures, multiple years have been combined to enhance the reliability of the estimates. In these cases, average annual rates have been calculated by dividing the multiple-year estimate by the sum of the area’s population count across the multiple years. For instance, for the death measures, the death counts for a five-year period (1992-1996) were divided by the population counts for 1992-1996 combined. The average annual number of events have been reported in the reference tables.

Two areas (#35, South Jordan and #46, East Orem) contain zip codes that were created recently (1993 and 1996, respectively). For measures that rely on combining data over multiple years, the estimates for those areas will be based on smaller populations (e.g., a population over one year instead of five). Because of the smaller population base, the precision of the estimates for areas #35 and #46 will not be as good as it would otherwise have been. In addition to lack of precision in the estimates, it is likely that use of new ZIP codes does not begin uniformly on the date the ZIP code change was initiated. It is very possible that some events that took place in areas #35 and #46 after creation of the new ZIP codes were improperly coded as having taken place in areas #39 (Riverton) and #45 (West Orem), respectively, the areas that once included the new ZIP codes. **Reported rates calculated for areas #35 and #46 should be interpreted with caution.**

Age-Adjustment

When comparing rates across geographic areas, the rates to be compared are typically age-adjusted to control for area-to-area differences in health events that can be explained by differing ages of the area populations. For example, an area that has an older population will have higher crude (not age-adjusted) rates for cancer, even though its exposure levels and cancer rates for specific age groups are the same as those of other areas. One might incorrectly attribute the high cancer rates to some characteristic of the area other than age. Age-adjusted rates control for age effects, allowing more meaningful comparisons of rates across areas.

The age-adjusted death rate is most often computed using the direct method, as it is the simplest and most straight-forward method of standardization. Direct standardization adjusts the age-specific rates observed in the small area to the age distribution of a standard population (Lilienfeld & Stolley, 1994). Using direct standardization, the age-adjusted death rate is a weighted average of the age-specific death rates, where the age-specific rates are the relative age distribution of the standard population (i.e., the percentage of the standard population in each age group).

Direct standardization can present problems when age-group-specific rates for small areas are unstable. In such cases, indirect standardization of rates may be used. Indirect standardization adjusts the overall standard population rate to the age distribution of the small area (Lilienfeld & Stolley, 1994). Indirectly standardized rates are based on the standard mortality or morbidity ratio (SMR) and the crude rate for a standard population.

An indirectly standardized death or disease rate (ISR) can be computed as:

$$ISR = SMR*R_s$$

$$SMR = \frac{\text{observed deaths/disease in the small area}}{\text{expected deaths/disease in the small area}} = \frac{D}{e} = \frac{D}{(R_{si} * P_i)}$$

Where...

R_s = the crude death/disease rate in the standard population

R_{si} = the age-specific death/disease rate in age group i of the standard population (# deaths/population count)

P_i = the population count in age group i of the small area

It is technically appropriate to compare indirectly standardized rates only with the rate in the standard population, not with each other.

Age and sex adjusted birth rates (ASABR) were computed using the following formula:

$$ASABR = \sum f_a (P^f_a / P) * 1000$$

Where...

$\sum f_a = \sum (b_a / P^f_a)$ is age specific birth rate in a particular population

P^f_a is the age-specific female population count in the standard population

P is the total population (both sexes) in standard population

Age-adjusting is not necessary when only age-specific rates are used, when the population of study has a narrow age range, or when it is desired to report the crude rate, regardless of age effects. Age-adjustment and calculation of 95% confidence intervals for the 1996 Utah Health Status Survey were accomplished using SUDAAN software (Shah, Barnwell & Bieler, 1997), which takes into account the design effects inherent in complex survey data (Lee, Forthoger & Lorimor, 1989).

Calculation of Confidence Limits

A rate calculation may be of limited value when derived from a small population. Rates based on small numbers are statistically more likely to be affected by chance variation and have large variability over time. One way of dealing with small numbers is to use confidence intervals for help in interpreting the rates. The confidence interval is a range of values within which the "true" value of the rate is expected to occur.

A common formula for calculating a confidence interval is that for a proportion: CI = 1.96 * {SQRT of [(p * 1-p)/n]} where p = the proportion, and n = the size of the population at risk. When calculating confidence intervals for rates based on rarely occurring events (fewer than 20 events), the formula differs, and the Poisson distribution must be used. The confidence interval for directly standardized rates (DSR) can be computed as follows:

$$\begin{aligned} CI(DSR) &= \pm 1.96 * SE(DSR) * K \\ &= \pm 1.96 * SQRT(VAR(DSR)) * K \\ &= \pm 1.96 * SQRT(W_i^2 * Var(R_i)) * K \\ &= \pm 1.96 * SQRT(W_i^2 * ((R_i * (1 - R_i))/P_i)) * K \end{aligned}$$

Where...

SE(DSR) = the standard error of the directly standardized rate
K = a constant (e.g., 100,000) that is being used to communicate the rate
W_{si}^2 = the population weight for the ith age group in the standard population
R_i = the age-specific death/disease rate in the ith age group of the small area population (# deaths/population count)
P_i = the population count in age group i of the small area

For indirectly standardized rates based on events that follow a Poisson distribution and for which the ratio of events to total population is small (<.3) and the sample size is large, the following two methods can be used to calculate confidence interval (Kahn & Sempos, 1989).

(1) When the number of events >20:

$$CI(ISR) = (SMR \pm 1.96\ SQRT(SMR/e)) * R_s * K$$

Where...

R_s = the crude death/disease rate in the standard population
K = a constant (e.g., 100,000) that is being used to communicate the rate
SMR = observed deaths in the small area / expected deaths in the small area
e = expected deaths/disease in the small area = $(R_{si} * P_i)$
R_i = the age-specific death/disease rate in the ith age group of the small area population (# deaths/population count)
P_i = the population count in age group i of the small area

(2) When the number of events <=20:

$$LL(ISR) = (\text{Lower limit for parameter estimate from Poisson table/e})) * R_s * K$$
$$UL(ISR) = (\text{Upper limit for parameter estimate from Poisson table/e})) * R_s * K$$

Where LL is the lower confidence interval limit, and UL is the upper confidence interval limit. For measures based on events that occurred over multiple years, the number of events refers to the total number of events during the time period, an not the average annual number of events.

Mapping and Statistical Software

ArcView geographic informations system software was used to create the maps found in this report. The software applications used for data analysis included SAS and SUDAAN.

Definition of Measures

Measure	Definition	Data Source[4]	Denominator	Age-Adjustment	Confidence Limits
Median Age	Weighted average of the median age of all ZIP code areas included in the small area.	CACI Marketing, Inc.	All members of the population	No	Not calculated
Per Capita Income	Weighted average of the per capita income of all ZIP code areas included in the small areas.	CACI Marketing, Inc.	All members of the population	No	Not calculated
Overall Health Status	Survey item[1]	1996 Utah Health Status Survey, Utah Department of Health	All members of the survey sample	Age-adjusted to 1996 Utah population using SUDAAN software	95% confidence limits calculated using SUDAAN software
Cigarette Smoking	Survey item[2]	1996 Utah Health Status Survey, Utah Department of Health	All members of the survey sample	Age-adjusted to 1996 Utah population using SUDAAN software	95% confidence limits calculated using SUDAAN software
Health Insurance Coverage	Survey item[3]	1996 Utah Health Status Survey, Utah Department of Health	All members of the survey sample	Age-adjusted to 1996 Utah population using SUDAAN software	95% confidence limits calculated using SUDAAN software
Birth Rate	Number of births per 1,000 population	UDOH Vital Records births data set	All members of the population	No	95% confidence intervals calculated (1.66 * s.e.)
General Fertility Rate	Number of births per 1,000 women age 15-44	UDOH Vital Records births data set	All females in the population age 15-44	No	95% confidence intervals calculated (1.66 * s.e.)
Births to Adolescents	Number of births to adolescents age 10-17 per 1,000 adolescent females age 10-17 in the population	UDOH Vital Records births data set	All females in the population age 10-17	No	95% confidence intervals calculated (1.66 * s.e.)

Definition of Measures (continued from previous page)

Measure	Definition	Data Source[4]	Denominator	Age-Adjustment	Confidence Limits
Low Birth Weight	Percentage of live-born infants weighing less than 2,500 grams at birth	UDOH Vital Records births data set	All live-born infants during the time period	No	95% confidence intervals calculated (1.66 * s.e.)
Prenatal Care	% of mothers delivering live infants who did not receive prenatal care in the first trimester.	UDOH Vital Records births data set	All live-born infants during the time period	No	95% confidence intervals calculated (1.66 * s.e.)
Infant Mortality	Deaths among infants under 1 year of age per 1,000 live births	UDOH Vital Records births and deaths data sets	All live-born infants during the time period	No	95% confidence intervals calculated (1.96 * s.e.)
Deaths from All Causes	Deaths per 100,000 population. ICD-9 codes 001-999	UDOH Vital Records deaths data set	All members of the population	Age-adjusted to Utah 1990 population using the indirect method[5]	95% confidence intervals calculated for indirectly standardized rates[6]
Motor Vehicle Crash Deaths	Motor vehicle crash deaths per 100,000 population. ICD-9 codes E810-E825	UDOH Vital Records deaths data set	All members of the population	Age-adjusted to Utah 1990 population using the indirect method[5]	95% confidence intervals calculated for indirectly standardized rates[6]
Suicide	Suicides per 100,000 population. ICD-9 codes E950-E959	UDOH Vital Records deaths data set	All members of the population	Age-adjusted to Utah 1990 population using the indirect method[5]	95% confidence intervals calculated for indirectly standardized rates[6]
Lung Cancer Deaths	Lung cancer deaths per 100,000 population. ICD-9 code 162	UDOH Vital Records deaths data set	All members of the population	Age-adjusted to Utah 1990 population using the indirect method[5]	95% confidence intervals calculated for indirectly standardized rates[6]
Cardiovascular Disase Deaths	Cardiovascular disease deaths per 100,000 population. ICD-9 codes 390-448	UDOH Vital Records deaths data set	All members of the population	Age-adjusted to Utah 1990 population using the indirect method[5]	95% confidence intervals calculated for indirectly standardized rates[6]

1. "In general, would you say your health is excellent, very good, good, fair, or poor." Survey items were reported by one survey respondent for all household members.
2. Has smoked at least 100 cigarettes in lifetime, and was a current smoker at the time of the survey. Survey items were reported by one survey respondent for all household members.
3. The next few questions ask about health insurance. By health insurance I mean private and employer plans, prepaid plans such as HMOs, and government plans, such as Medicare. Are all, some, or none of the members of your household currently covered by health insurance? [If "some"] Which members of your household ARE covered by any kind of health insurance, public or private? Survey items were reported by one survey respondent for all household members.
4. All population estimates were purchased from CACI Marketing, Inc.
5. The indirect method of age-adjustment was used because there were small numbers of deaths in individual age strata. Indirect standardization adjusts the overall standard population rate to the age distribution of the small area (Lilienfeld & Stolley, 1994). It is technically appropriate to compare indirectly standardized rates only with the rate in the standard population, not with each other.
6. Confidence intervals were calculated using a method recommended for indirectly standardized rates (Kahn & Sempos, 1989). For the death data, 95% confidence intervals were used.