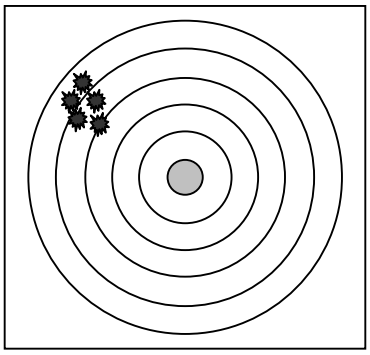
Appendix G: Statistical Reliability and Validity

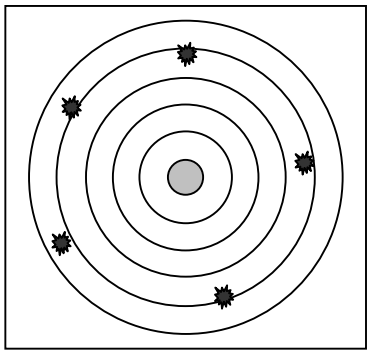
We use measures of health status in an attempt to understand the underlying disease risk in a population. For instance, if a certain city has a consistently high rate of food poisoning, we would want to investigate the food establishments in that city in an attempt to reduce the risk of food poisoning. This report is an exercise in surveillance among Utah's race and ethnic communities. If one community has a consistently high rate of a certain disease, we would want to investigate it further in an attempt to identify and reduce the disease risk in that community.

In practice, public health surveillance uses objective measures, such as rates of death, illness, injury, and hospitalization to indicate a potential problem, one that might merit further investigation. Many objective measures have been presented in this report. To successfully interpret the measures in this report, we need to know something about how well the measure represents the underlying disease risk in the community. There are two important elements involved in the quality of a measure: reliability and validity.

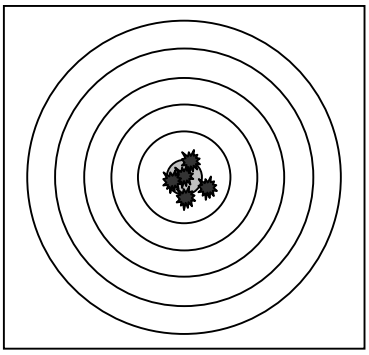
In the three figures, below, the bull's-eye of the target represents the true underlying risk of disease in a population, and the holes in the target represent multiple objective measurements of the risk. In the first figure, the measure is reliable—it measures nearly the same value each time. But the measure in figure 1 is not valid—the average of the scores is not close to the true underlying risk. In the second figure, the scores are not very reliable—there is a lot of variability in the scores, but they center around the true risk value. In the third figure, the measure is both reliable and valid. The term "accuracy" is often used in relation to validity, while the term, "precision" is used to describe reliability.



1. Good reliability, poor validity.



2. Poor reliability, good validity.



3. Good reliability, good validity.

In public health, we are quite lucky that the validity of our measures is really quite good. Cause of death on death certificates is certified by a physician. Survey measures have been tested to maximize validity. Birth weight is reported at the birth hospital. There are some issues with the validity of the measurement of race and ethnic group status (discussed in Appendix H), but on the whole, the measures we use have a high degree of validity.

The underlying population risk for a given health problem will be relatively stable, but our measures of the problem itself will have variability, even when the measurement is drawn from the entire population. That variability indicates poor measure reliability. The reasons for the variability include primarily three factors: (1) the health events are relatively rare, (2) the population size is relatively small, and (3)

the health events do not occur at regularly occurring intervals. For instance, infant mortality is an extremely important indicator of health status and access to care in a given population. But it is relatively rare—occurring in only about 5 out of 1,000 births. Measured across all births in Utah, the measure is fairly reliable (5.2 ± 0.3 infant deaths per 1,000 births between 1998 and 2003). In Utah's Black/African American community, however, the infant mortality rate over the same time period (13.8 infant deaths per 1,000 births to Black/African American women) had a 95% confidence interval of ± 5.3. The measure, infant deaths, has virtually the same validity in the Black/African American population as it does in the overall state population. But because infant deaths are relatively rare, the population of Black/African American women giving birth is relatively small, and infant deaths do not occur at regularly-timed intervals, the time sample we have used (1998–2003) produces a measure that is less precise in the Black/African American population than it is in the entire state.