Data Issues in Areas with Small Populations

Data collection is an integral part of the assessment function of public health. The challenge is to collect and convert data into useful information that provides a composite picture of the community's health.

Sparsely populated or small communities have the unique challenge of translating "low numbers or incidence" into usable information. Low numbers or incidence can produce unstable rates that greatly fluctuate from year to year. In addition, a "snapshot" view of one year may not adequately represent the true status of the community's health. Also, smaller communities may not have access to individuals with expertise in data analysis. For these reasons, the collection and analysis of data may be an especially large barrier to community health assessment in communities with small populations.

Statistical Instability

There are two potential ways to avoid or address the statistical instability with which jurisdictions with low populations are faced. It is recommended that such communities consider one of the following approaches:

- Combine multi-year data (e.g., data for three years). A drawback to this option is that looking at multi-year data limits the ability of the jurisdiction to monitor program interventions and identify new trends. Rolling year averages (e.g., looking at data for 1997-2000 one year, and 1998-2001 the following year) may overcome this drawback and should be considered.
- Expand the geographic area by conducting a regional health assessment in collaboration with neighboring jurisdictions. A drawback to this option is that the community may then be looking at geographical areas over which it has no control. Analyzing data at the regional level may also mask interesting local variations in the data.

Both of these approaches increase the number of events under analysis. It is recommended that all indicators be based on 20 or more events (i.e., infant deaths, low birth-weight infants, etc.) In general, the higher the number of events, the more stable is the data. Confidentiality issues must also be considered when the number of events is small.

Other Data Considerations

The following tips may be useful to communities that do not have access to epidemiological expertise in data analysis.

Data should be considered in light of the following questions: What are the sources of these data? Are the sources reliable? What are the issues raised by the data? Are key pieces of information missing and can they be obtained? Are there any other considerations regarding the health issue that need to be taken into account when analyzing the data? Can a summary statement be made about the numbers?

Consider the following issues:

- The manner in which the data are collected is very important. In analyzing communicable disease information, consider the reporting system that exists in the state or locality and the kind of data that would be produced.
- Consider the sampling frame used in gathering the data to ensure that all special high-risk populations are included. For example, university populations should be

included when looking at youth issues, while nursing home or retirement communities should be included to get a valid rate for aging issues.

- Consider time-related issues when looking at certain health issues or diseases (i.e., the amount of time it may take for a program intervention to show results). For example, a decrease in cancer rates may be indicative of the success of program interventions that took place many years earlier.
- Years of Potential Life Lost (YPLL) is a good indicator that can provide additional information about the important causes of premature death in a community. For example, consider the number of deaths due to injuries in a community. Although the actual number of deaths due to injuries might be low, the impact of this problem could be highlighted if the YPLL is high (indicating that deaths due to injuries cause a disproportionate loss of potential productivity in younger populations).
- Consider that a substantial change in a single indicator (e.g., number of cancer deaths increasing from 20 to 30 [or a 50% rise] over one year) may not necessarily represent a trend or pattern. While troubling to the community, this may be a normal variation in reporting. Situations like this may present an opportunity to engage the community in the science of epidemiology. Exploring risk factors may increase the participants' appreciation for health planning, community health status assessments, and related activities.

References:

American Public Health Association. Healthy Communities 2000: Model Standards. Washington, DC: 1991, pp. 458-459.