The Issue of Multiple Races in Calculating Disease Rates
Part one of a two-part series.

Background.

Accurate, consistent and comparable data are important to the ability of the Board and its member tribes to understand health-related risk and disease burden in Northwest tribal communities. In addition, because disease rates and the characterization of disease within a population are often used in the allocation of funding for, and the planning and management of public health programs, it is especially important that accurate and reliable data are available.

In order to evaluate or calculate rates for any disease, it is important to remember exactly what a rate is. A rate is a statistical term that tells us about the force of a disease in a population. It includes the number of new cases of disease in the numerator (the part of a fraction above the line), and persons and time in the denominator (the part of a fraction below the line). For example, let's suppose that there were 100 new cases of cancer among 100,000 people over the course of one year. The rate of cancer would look like this-

$$\frac{\textbf{100 cancer cases}}{\textbf{100,000 persons at risk per year}}$$

OR

100 per 100,000 person-years

When calculating a rate, there are several important things to keep in mind. First, rates requires that only people who are at risk for the disease can be counted in the denominator, and that only people who are included in the denominator, can be counted in the numerator. Simply stated, the cases in the numerator must be among people in the denominator. Second, only rates that are calculated using exactly identical methodology can be compared. Although this may appear relatively straightforward, when dealing with race-specific rates, problems can arise.

The Problem.

The data source most commonly used as the denominator for calculating disease rates is the US Census. The Census provides an enumeration of the entire US population, and includes important demographic information like gender, age, race and ethnicity.

The 2000 Census allowed people to self-identify as having more than one race in their ancestry. For example, an individual whose mother is American Indian, and whose father is white, was able to select both "AI/AN" *and* "white" on the 2000 US Census form. Prior US Censuses (as recent as 1990), and almost all state and federal data systems from which health-related information is derived (including disease registries) have traditionally allowed for the reporting of only one race. So using the example above, the individual would have only been able to select only one race, either “AI/AN” *or* “white”. Thus, calculating race-specific disease rates is problematic because the numerator data, (often collected from state and federal data systems that use single-choice race categories) are incompatible with Census-derived denominators (data collected by the 2000 US Census using multiple-race categories). Using the same example, an individual who is AI/AN and white might be included in data used to calculate disease rates for the AI/AN population only, the white population only, both the AI/AN and white populations, or neither. The impact of these different allocations is profound. A rate can increase or decrease dramatically based on what data are used to calculate the rate, ultimately providing very different estimates of disease burden within a population. This problem represents an important obstacle to efforts by the NPAIHB to characterize the health status of AI/ANs living in the Northwest.

The Solution?

In order to assess the health status of Northwest AI/ANs and continue working toward the goal of reducing racial health disparities, local, state, regional and federal health agencies need to adopt a standard approach to dealing with the issues of multiple races and the calculation of race-specific disease rates.

The National Center for Health Statistics, in collaboration with the US Census Bureau recently released “bridged” population estimates based on the 2000 Census Data. The bridged data results from a proportion (derived from the 1997-2000 National Health Interview Survey) being applied to multiple race categories, in order to convert them to single-race categories. Using the original example, in the bridged estimates, an individual who is both AI/AN and white is split, metaphorically, so that a portion of the person (say 2/3) is counted in the AI/AN population, and the remainder (1/3) is counted in the white population. Using a bridged methodology helps to eliminate the problem of overestimating the population (which occurs if every individual who selected multiple races was counted in every selected race group) and the problem of underestimating certain populations (which occurs if every individual who selected multiple races was counted in only one of the selected race groups).

Ultimately, bridging multiple race categories appears to be a useful and logical, though imperfect, way to address the issue of multiple races and calculating accurate race-specific disease rates. It is important to note that the bridging process does introduce its own set of errors into population estimates, and subsequently into the disease rates based upon those estimates. The potential for error is greatest for small populations and race-groups, though efforts are being taken to identify and address these errors. The effect of bridging population estimates and racial misclassification on disease rates, will be discussed in the second part of this series to be included in the next issue of Health News and Notes.

For additional information, please contact:
Emily Puukka, MS
Project Director, Northwest Tribal Registry Project
503.228.4185 Ext. 285
epuukka@npaihb.org