



## **DATA: IT'S LIMITATIONS AND HOW IT WAS COLLECTED.**

1. We relied on existing data from several sources in the community, e.g. the school, the Health Dept., law enforcement, etc. We also used state and national data sources. For example, the N.C. Child Advocacy Institute provides some data for both Henderson County and North Carolina.
2. Ideal data – is consistent over time and across jurisdictions so trends and comparisons can be made. However, that is not how it is in the real world.
3. Limitations of data. Availability, how things are counted, how rates are determined, and comparability

- Availability – This means the data are not available at all or not in the form we would like to have. For example, we may not be able to get trend data. Also, community professionals may see a problem for which no data are collected or the data that are collected are inadequate. We ran into this trying to get data on truancy and children's dental needs as examples.
- How things are counted – In order to compare data over time or between jurisdictions, the definition of the measurement must stay the same. For example, the definition of “drop-out” has not been consistent over time or between jurisdictions. It is also important to understand how things are measured or the definition of a term such as “drop out.” What does it include? What does it not include? We tried to include the definitions of the measures in Appendix A whenever they were available. Having the definitions of the measures provides a better understanding of the data.
- How rates are determined – Data on people, such as birth rates, are significant only if considered in relation to the size and kind of population from which they are drawn. It is important to present the number of cases and the rates. Rates standardize the data across time and across jurisdictions based on differences in the sizes of the populations.

Sometimes rates are given but not the raw numbers. For example, data may be presented in percentages or rates per 1,000. One statistic we saw gave the cumulative rate per 1,000 over a period of years, which is almost impossible to use for comparisons. In other situations, the number of cases may be given but not the rates.

Rates are based on the (numerator) number of cases divided by a (denominator) size of the population times 100 for the percentage and 1,000 for the rate per 1,000. When the rate is not specified, it is difficult to determine what it is because the size of the relevant population may not be easily available. For example, to determine the rate of teenage pregnancies for females ages 15-19 in 1999 in Henderson County, you have to have the number of females in that age group for that year in the county.

The major source of population data is from the Census Bureau. The decennial census is only collected every ten years, so population characteristics have to be estimated between decennial censuses. The data may not be available in enough detail at the county level or estimates may be flawed.

Each state has a State Data Center, which is the resource for all data the Census Bureau collects. The Census Bureau conducts several surveys between decennial censuses, such as the Current Population Surveys. These are sample surveys rather than censuses. They are used to make estimates of population characteristics between censuses.

- Comparability – If comparable measures are used across time and between jurisdictions, it is possible to look at trends and compare Henderson County with North Carolina and the nation.

Given the limitations of the data, how can it be used to identify potential problems? The first thing is to understand the data.....what was the measurement? Did it change? How were the rates determined, etc.?

*Continued*

Then you triangulate. That means that you look at groups of what we call “indicators” rather than rely totally on one set of data.

Input from community professionals is also important and needed. Sometimes they see things that help us interpret the data or identify problems that the data are missing.

### **Criteria for Identifying Best Measures**

There are three criteria which can be used to identify the best measures:

**Communication Power:** Does the indicator communicate to a broad range of audiences? It is possible to think of this in terms of the public square test. If you had to stand in a public square and explain to your neighbors “what we mean, in this community, by children healthy and ready for school,” what two or three pieces of data would you use? Obviously you could bring a thick report to the square and begin a long recitation, but the crowd would thin quickly. It is hard for people to listen to, absorb or understand more than a few pieces of data at a time. They must be common sense, and compelling, not arcane and bureaucratic. Communication power means that the data must have clarity with diverse audiences.

**Proxy Power:** Does the indicator say something of central importance about the result? (Or is it peripheral?) Can this measure stand as a proxy for the plain English statement of well-being? What pieces of data really get at the heart of the matter?

Another simple truth about indicators is that they run in herds. If one indicator is going in the right direction, often others are as well. You do not need 20 indicators telling you the same thing. Pick the indicators which have the greatest proxy power, i.e. those which are most likely to match the direction of the other indicators in the herd

**Data Power:** Do we have quality data on a timely basis? We need data which is reliable and consistent. And we need timely data so we can see progress—or the lack thereof—on a regular and frequent basis. Problems with data availability, quality or timeliness can be addressed as part of the data development agenda