

## *The Rural Early Childhood Information Gap*

---

The shift of public policymaking regarding children and families from the federal to the state and local levels has increased demand for measures of child well-being in states and communities. However, it is difficult to obtain reliable data about the well-being of rural children in states or communities. Of 61 key indicators of child well-being, 51 cannot be estimated for rural children using public-use data and a precise definition of rurality. Even using non-public-use data, most of these indicators can only be estimated for rural children under 18 nationally and in some larger states.

Two primary issues impede or preclude calculation of many well-being indicators for rural children: the rules governing public release of data which might be used to identify individual respondents, which are designed to protect the confidentiality of respondents, and the size of the sample of rural respondents.

### **Data Confidentiality Rules**

Survey research firms and agencies that collect data on human subjects must protect the privacy of the individuals from whom data are collected. Strict data dissemination rules control the information that is released to the public, particularly information that could play a role in identifying individual respondents. The obvious personal identifiers — names, Social Security numbers and addresses — are not released. In addition, variables that place respondents in small geographic areas, such as neighborhood, city, county, or even state, often are suppressed because these variables could be used in conjunction with other variables to identify specific respondents. Thus, variables

that simply identify whether a respondent lives in a rural area are often unavailable.

The U.S. Census Bureau and most other agencies that collect data on human subjects do not release variables that allow researchers to identify respondents in areas with a population of less than 100,000 residents. Many major public-use datasets only indicate whether a respondent lives within or outside a metropolitan statistical area (MSA). The major data sources administered by the U.S. Census Bureau — the Current Population Survey (CPS), Census 2000, and the Survey of Income and Program Participation (SIPP), as well as the National Vital Statistics System (NVSS) — only release this less precise measure of rurality.

The inside/outside MSA distinction is a poor measure of rurality because many respondents who live in rural areas (with population of less than 2,500) are incorrectly coded as metropolitan, and vice versa. Moreover, the lack of a consistent definition of rurality often complicates discussions of rural issues. This is true not only in the research literature, but also in the data collection that provides the underlying information for analysis.

Most organizations and agencies that maintain non-public datasets with information about rurality do allow researchers to use the datasets, but the process can be burdensome and costly. For example, using non-public-use data of the U.S. Census Bureau can cost more than \$35,000 per year, and there are numerous limitations on the types of results that can be reported from the non-public-use data.

## Sample Size

Generally, the more detailed a survey is, the fewer the number of individuals surveyed. This means that while national datasets containing larger numbers of indicators of child well-being may be used to estimate national averages for rural young children, they do not involve enough individual respondents to be reliable at the state or community level.

The comparison of Census 2000 data and the 2003 Current Population Survey provides an excellent example of this phenomenon. Sample size is not an issue when using the Census 2000 data, even at the lowest levels of geography. However, this data source only contains basic demographic information on children. The Current Population Survey, in contrast, contains a wealth of indicator information, but contains significantly smaller samples. As such, the indicators derived from Currently Population Survey data cannot be applied to lower levels of geography in most cases.

It is important to note that administrative data sources like the National Vital Statistics System or Adoption and Foster Care Analysis and Reporting System have a record of every occurrence of the event, behavior, or activity that is measured in the data file, so these sources do not have the same limitations of sample size as do probability samples.

## Conclusion

More should be done to promote the analysis and dissemination of child well-being indicators for rural children. Indeed, with millions of children living in rural areas across the country, more must be done to ensure that public policy anticipates the needs of these children. This is especially important given the extra challenges that long distances pose for services to rural children and their families, and given the greater role that state and local governments play in child and family policy.

This Rural Early Childhood Brief is a summary of a lengthier Rural Early Childhood Report, "Young Children and the Rural Information Gap: The Weaknesses of Major Data Sources for Examining The Well-being of Rural Children," by Jeffrey Capizzano and Alexandra Fiorillo. See the full report at <http://www.ruralec.msstate.edu/reports/>.

© 2004 National Center for Rural Early Childhood Learning Initiatives

P.O. Box 6013  
Mississippi State, MS 39762  
[www.ruralec.msstate.edu](http://www.ruralec.msstate.edu)