Identifying and Analyzing Geographic Change to School Districts

Colleen D. Joyce
U.S. Census Bureau
Housing and Household Economic Statistics Division
Room 1451-3
Washington, DC 20233-8500
Colleen.D.Joyce@census.gov

The U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program produces poverty and income estimates for states, counties, and school districts on an annual basis. These estimates provide updated income and poverty statistics, which are used for the administration of federal programs and the allocation of federal funds to local entities.

Although SAIPE's main reason for producing the estimates is to provide the U. S. Department of Education with the necessary information to allocate Title I funding under the No Child Left Behind Act of 2001, the estimates are used by a variety of data users for a variety of purposes. Some data users use the annual data stand-alone, but others are interested in using the annual estimates to explore how poverty and income has changed over time.

SAIPE's goal is to produce the best estimate possible for a specific point in time. The estimates are not intended to be used in time series analyses. However, should data users choose to analyze the estimates in a time series, it is important they be made aware of the caveats involved with doing so.

When a change in the estimate for a specific entity is observed from one estimate year to another, a number of reasons might explain it. These reasons can be roughly categorized into three groups: those involving geographic change, those involving universe change, and those with estimated demographic change. In many cases, the demographic change is what data users are really interested in. However, even when data users can isolate demographic change from geographic and universe changes, there are still numerous

issues involved with comparing SAIPE data for the same area across years. These issues have been documented by the SAIPE team, and are outlined on SAIPE's website.[1] Less well documented are geographic and universe change issues. This paper will focus primarily on these two issues, and specifically on how these types of changes are accounted for in the estimates and how the impact of these changes can be determined. Because there is little change in the geography and universe at the state or county level, the paper will focus primarily on the school district estimates.

How the Estimates are Created

Before looking at the issues associated with analyzing the estimates, it is necessary to have a basic understanding of how the estimates are created.

For state and counties, estimates are released for:

- the total number of people in poverty;
- the number of children under age 5 in poverty (for states only);
- the number of related children age 5 to 17 in families in poverty;
- the number of children under age 18 in poverty; and
- median household income

In addition, SAIPE produces the following for school districts eligible for Title I funding under the No Child Left Behind Act of 2001:

- the total population;
- the number of relevant children age 5 to 17; and
- the number of related, relevant children age 5 to 17 in families in poverty

[1] Detailed documentation regarding uncertainty in the estimates and cautions associated with comparing modeled estimates of the same county in different years can be found on SAIPE's website at http://www.census.gov/hhes/www/saipe/

Relevant children or population refers to the children or population served by the school district. For example, the relevant children for an elementary school district that serves kindergarten through grade 8 would be the population age 5-13. For a secondary school district serving grades 9-12, the relevant population would be that population age 14-17. A unified district serving grades K-12 would have a relevant population equal to that population age 5-17. Figure 1 shows the location of elementary, secondary and unified districts.

State and County Estimates

The poverty and income estimates start with national estimates obtained through the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). State and county estimates are created using a model-based approach. Inputs to the model include the CPS ASEC data, and other tax and program data such as:

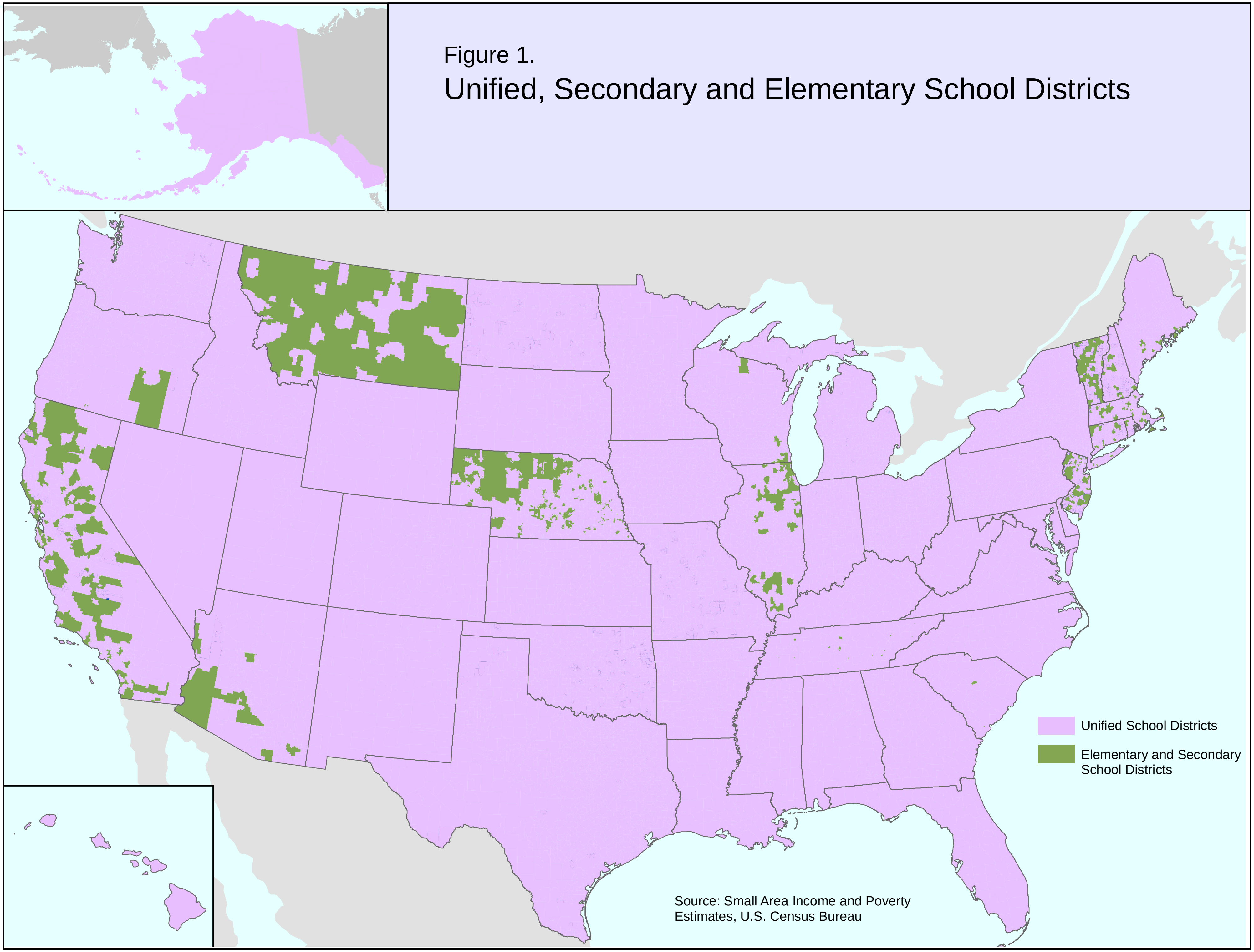
- Internal Revenue Service (IRS) tax return data
- counts of food stamp participants
- Bureau of Economic Analysis (BEA) income data
- decennial census estimates
- intercensal population estimates

School District Estimates

Much of the SAIPE models' input data cannot be uniformly geocoded to geography below that of the county level. It is for this reason that school district poverty estimates are arrived at using a different methodology. Once the estimate for the number of poor children in families in the county has been established, the relevant population is distributed among the school districts in the county. If a school district crosses the county line and is located in more than one county, the county population is distributed only to the piece of the district within the county.

Figure 1.

Unified, Secondary and Elementary School Districts



The distribution is made using the same proportions that existed in the decennial census. For example, suppose the decennial census estimated 100 poor children in county A, with 50 of those living in district one (50 percent), 25 in district two (25 percent), and 25 in district three (25 percent). The 2002 county estimated number of poor children, as determined by the model, is 200. 100 of those would be assigned to district one (50 percent), 50 to district two (25 percent), and 50 in district three (25 percent). (See Table 1.) That of course, is assuming that the school district geography has not changed since the decennial census. But what if the geography has changed?

Table 1. Distributing a County's Estimated Number of Relevant Children in Poverty Among School Districts Within that County

Geographic Entity	**Census 2000 number of relevant children age 5 to 17 in poverty**	**Census 2000 distribution of county's relevant children in poverty to school districts**	**2002 estimated number of relevant children age 5 to 17 in poverty (assuming no geographic changes)**
County A	100	-------	200
School District One	50	50%	100
School District Two	25	25%	50
School District Three	25	25%	50

Accounting for geographic change at the state and county level

Although rare, should a geographic change occur in any state or county boundary, that change would be accounted for in the models through the input data. IRS data, BEA income data, and food stamp data would be geocoded to the updated geography. Decennial census estimates are retabulated to the new geography through the Geographic Update System to Support Intercensal Estimates (GUSSIE).[2]

Accounting for geographic change at the school district level

GUSSIE retabulations are also used to create updated distributions of the number of poor children in whole school districts and school district pieces within counties. Building on the earlier example illustrated in Table 1, now assume that the boundary between school districts two and three has shifted. The original Census 2000 data showed that 50 percent of the poor children in County A were in district one, 25 percent were in district two, and 25 percent were in district three. After GUSSIE processes the boundary change between school districts two and three, the retabulated Census 2000 data shows that of the 100 poor children in the county, 50 of those are living in district one (50 percent), but now only 10 are in district two (10 percent), and 40 are in district three (40 percent). Again assume that SAIPE estimates 200 poor kids in the county in 2002. Based on the Census 2000 retabulation, 100 of those kids will be assigned to school district one (50 percent), 20 to district two (10 percent), and 80 to district three (40 percent). (See Table 2.)

Table 2. Example of How the Distribution of a County's Estimated Number of Relevant Children in Poverty is Distributed Among School Districts Within that County After Geographic Change

Geographic Entity	Census 2000 estimated number of relevant children age 5 to 17 in poverty	Census 2000 distribution of county's relevant children in poverty to school districts	2002 Retabulation of Census 2000 estimated number of relevant children in poverty (after boundary change between districts two and three)	2002 Estimate of number of relevant children age 5 to 17 in poverty (after boundary change between districts two and three)
County A	100	-------	100	200
School District One	50	50%	50	100
School District Two	25	25%	40	80
School District Three	25	25%	10	20

[2] See Appendix A for a description of GUSSIE.

If a data user were to look at the estimate of relevant children in poverty in school district two in 2000 and 2002, he or she would see that the number of poor children in the district increases from 25 to 80. What might not occur to the user initially is that part of that increase may not be due to demographic change, but simply to the fact that the district itself is larger, and encompasses population that was previously counted in another district.

So how can data users get a sense for how much of a given change in the estimates is due to geographic change and how much of it is demographic change? Examining the income year 2001 and 2002 poverty estimates might help to illustrate.

The 2001 and 2002 Poverty Estimates

In December 2004, SAIPE released income year 2001 and 2002 poverty estimates for school districts. Two years of data were released as the SAIPE program transitioned from a biennial to an annual release of data.

School District Boundary Collection and Differences Between Income Year and Boundary Year

Perhaps the first thing that data users should be aware of, is that the estimates for a specific income year do not always correspond with the boundary vintage year. (See Table 3.) Both the 2001 and 2002 estimates were based on school district boundaries as reported for the 2003-04 school year. The Census Bureau collected these school district boundaries in the fall of 2003. The Census Bureau contacts state officials every other year, giving them the opportunity to review the Census Bureau's school district information and provide updates and corrections to school district names, boundaries, and the grade ranges they serve.

Because these changes to school districts are only processed every other year, it is not always possible for the income year to match the school district boundary year. While

the income year 2002 estimates released in December 2004 are based on boundaries of a different year (2003-04), the income year 2003 estimates, scheduled for release in late 2005, will also be based on the 03-04 boundaries. The decision was made to use the most recent boundaries (03-04) for the 2001 and 2002 estimates (rather than the 01-02 boundaries), because it allows for more accurate allocation of funds under the No Child Left Behind Act of 2001.

Table 3. Relationship of Estimates, Boundaries and Data Releases

Estimates Income Year	School District Boundary Year	Year of Estimate Release
2002	2003-04	2004
2003	2003-04	2005
2004	2005-06	2006

Retabulating the Decennial Census

School District updates reported to the Census Bureau are processed through GUSSIE. During GUSSIE processing, Census 2000 data, including total population, population age 5 to 17, relevant population, relevant population in poverty, and housing unit counts, are retabulated to reflect updated boundaries and grade range assignments. The retabulated counts, referred to as the "base" counts, serve as inputs to the production of population and poverty estimates.

Because the base counts are a retabulation of the decennial census counts, and because the total counts from the census will not change, any changes in the total school district population base count from one year to the next will almost always be a result of geographic change. The Census 2000 total counts do not change, but the counted are now being assigned to different geography. Likewise, if the total base count for the area does not change but the population of relevant children does, a change in the grade range assignments, or universe, may be the cause. Analysis of the base counts allows us to isolate these geographic changes and analyze the implications of each on the estimates.

It should be noted that there are cases where changes in the population base counts result from circumstances other than changes in the boundaries. The Census Bureau is continuously improving our geographic databases. New and more accurate geographic information may show that we are geocoding housing units or group quarters to the wrong geography. Correcting the geocoding can result in units being "moved" to different geography without an actual change in the boundaries having occurred.

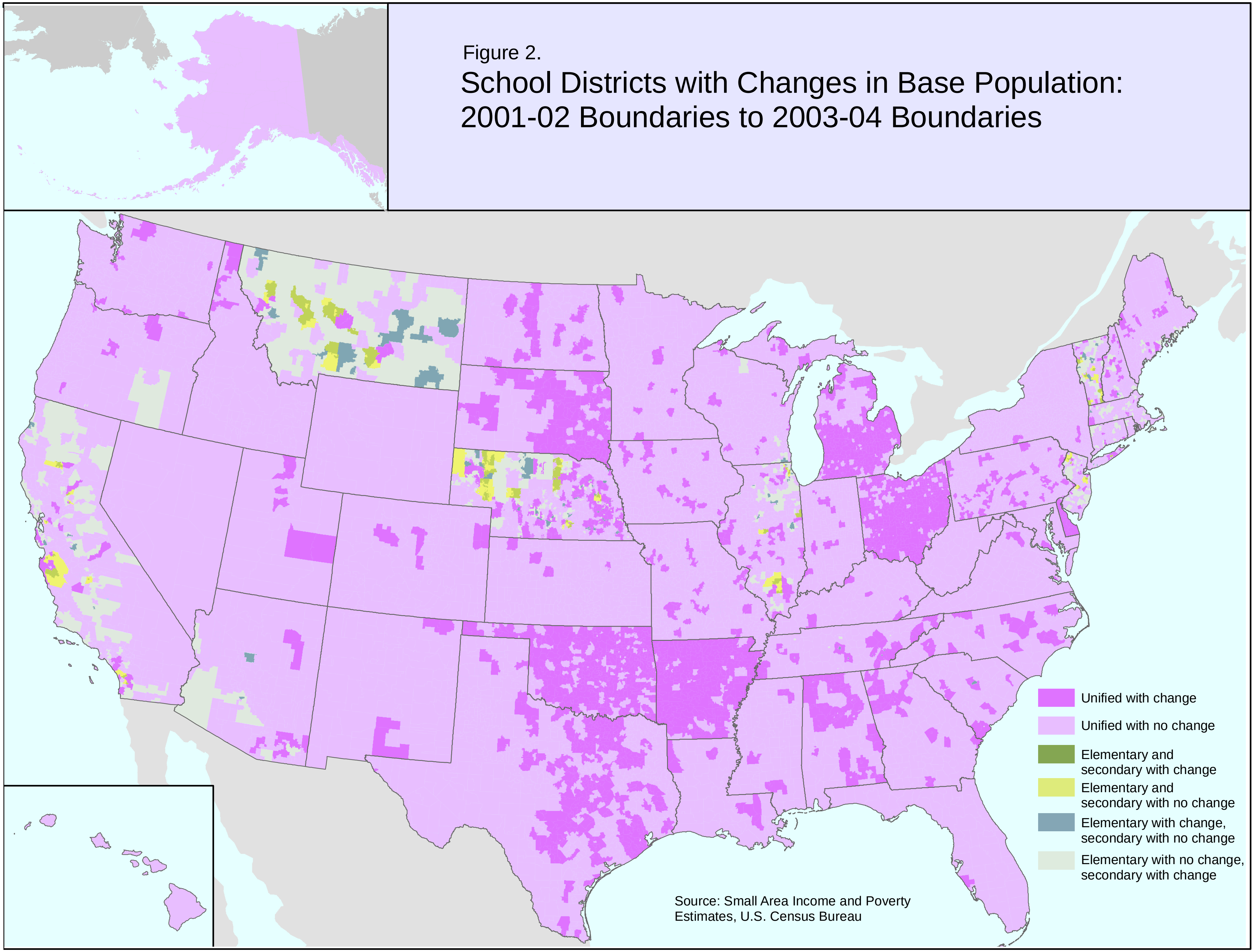
Of course, we do not need to look at the base counts to determine which districts had boundary or grade range assignment changes, since these changes are reported directly to us by state officials. However, looking at changes in the base counts can be extremely useful in determining the degree to which those changes affected the estimated populations.

How Many Districts Experienced These Types of Changes?

Comparisons between the school district total population base counts retabulated for the 2001-02 school district boundaries and those retabulated for the 2003-04 boundaries reveal that 3,238 out of 14,232[3] school districts, or 22.2 percent, experienced some base population change. (See Figure 2.) Net base count gains for districts ranged from 1 to 40,083 people. Net losses ranged from 1 to 29,927 people. New districts with as many as 16,199 people were created and districts with as many as 23,553 people were dissolved. Table 4 shows the number of districts with changes, broken down by the amount of change, and illustrates that the magnitude of change can vary widely. Of all school districts with changes, 25.1 percent involved net base population changes of 5 people or less. 53.5 percent involved 25 people or less, and over 25 percent involved changes of over 100 people. Figure 3 shows those school districts with numeric change and classifies the magnitude of that change.

Figure 2.

School Districts with Changes in Base Population: 2001-02 Boundaries to 2003-04 Boundaries



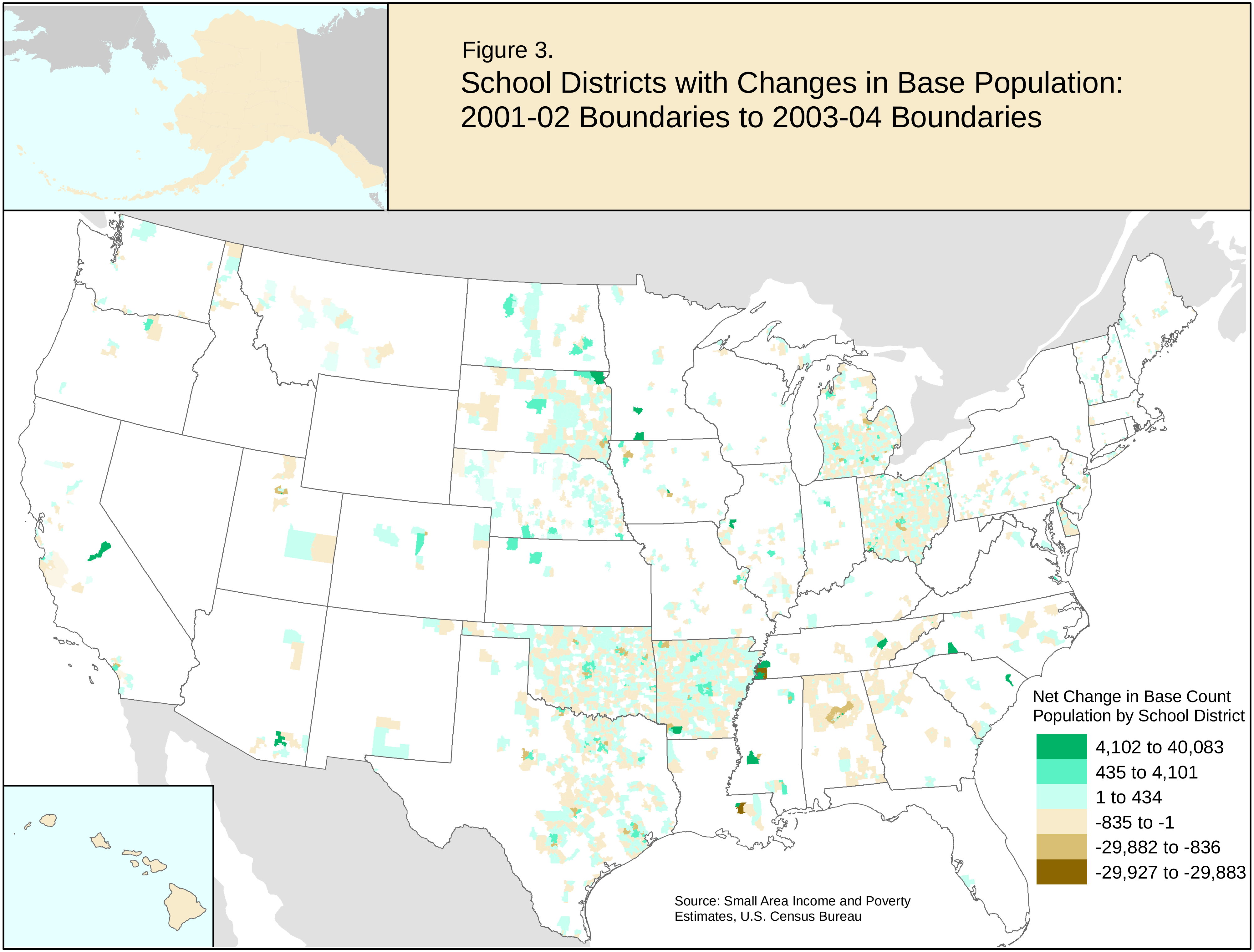


Figure 3.
School Districts with Changes in Base Population:
2001-02 Boundaries to 2003-04 Boundaries
Net Change in Base Count
Population by School District
4,102 to 40,083
435 to 4,101
1 to 434
-835 to -1
-29,882 to -836
-29,927 to -29,883
Source: Small Area Income and Poverty
Estimates, U.S. Census Bureau

Table 4. Number of School Districts with Net Numeric Base Population Changes

	Total Number of Districts	Net Numeric Change in Base Population							
		1-5	6-10	11-25	26-50	51-100	101-500	501-1000	1000+
Number of School Districts with Base Population Gains	1,629	396	206	260	175	170	284	60	78
Number of School Districts with Base Population Losses	1,609	417	196	259	166	159	255	69	88
Number of School Districts with Any Change (gains or losses)	3,238	813	402	519	341	329	539	129	166
Total Number of School Districts	14,232	*	*	*	*	*	*	*	*
Percent of all School Districts	22.8	5.7	2.8	3.6	2.4	2.3	3.8	0.9	1.2
Percent of all School Districts with Change	100.0	25.1	12.4	16.0	10.5	10.2	16.6	4.0	5.1

While Table 4 shows the number of districts with changes in the total base population, broken down by the amount of numeric change, Figure 4 and Table 5 shows the same districts, classifying them by the percentage change in their base counts. In this table, we can see that 69.1 percent of the districts with base population change had changes of less than 1 percent. 8.3 percent had changes of 11 percent or more.

[3] The total number of school districts (14,232) includes all school districts that existed based on the 2003-04 boundaries as well as school districts that were "dissolved" between the 2001-02 and 2003-04 boundary collections.

Figure 4.

School Districts with Changes in Base Population: 2001-02 Boundaries to 2003-04 Boundaries

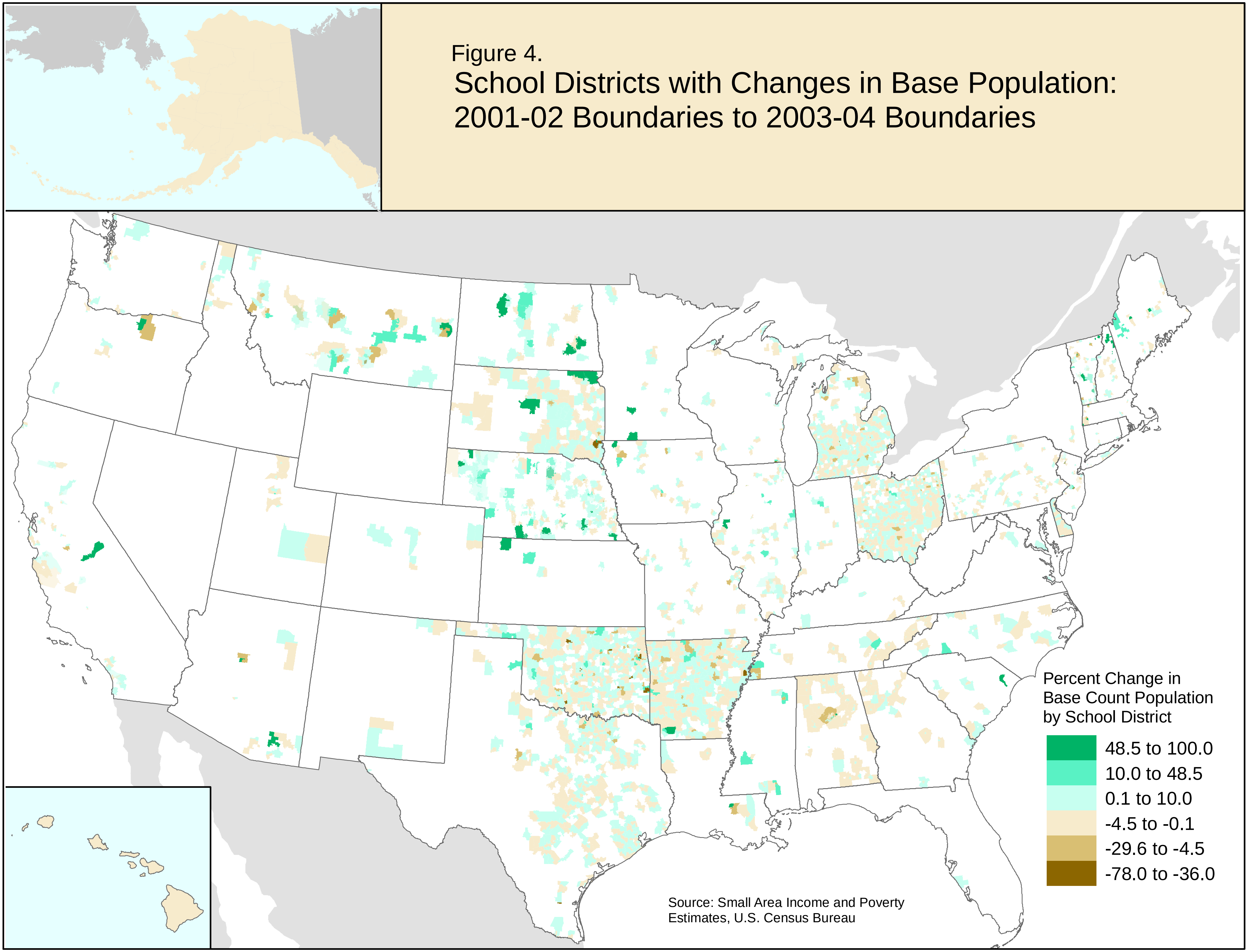


Table 5. Number of School Districts with Net Percent Base Population Changes

	Total Number of Districts	Percent Change in Base Population					
		Less than One	1.0 – 10.9	11.0 – 25.9	26.0 – 50.9	51.0 – 75.9	76.0 – 100.0
Number of School Districts with Base Population Gains	1,629	1126	386	43	29	0	42
Number of School Districts with Base Population Losses	1,609	1113	343	26	14	3	152
Number of School Districts with Any Change (gains or losses)	3,238	2239	729	72	43	3	152
Total Number of School Districts	14,232	*	*	*	*	*	*
Percent of all School Districts	22.8	15.7	5.1	0.5	0.3	0.0	1.1
Percent of all School Districts with Change	100.0	69.1	22.5	2.2	1.3	0.1	4.7

Case Study: Wheatland J1 Elementary School, Kenosha County, Wisconsin

Looking at a case study may help to better illustrate the impact of geographic change on the estimates and how the base population counts can inform the data user. Income year 1999 estimates, released in 2002, were based on 2001-02 school district boundaries. Poor children within Kenosha county were assigned to school districts using the same distribution that existed in Census 2000.

Because there was no boundary change reported for Wheatland J1 for the 2001-02 school year, the income year 1999 total population estimate and the Census 2000 data are the same. The income year 1999 estimated total population base count and the Census 2000 total population count for Wheatland J1 Elementary school district in Kenosha County, Wisconsin, was 2,703. The 2002 income year estimated total population (based on 2003-04 boundaries) was 4,213, a net increase of 1,510 total population, or 55.9 percent over the 1999 estimate and Census 2000 count.

When boundary updates were collected for the 2003-04 school year, Wheatland J1 reported an update that netted approximately 8.5 square miles for the district. Part of the net increase in total population and the population of poor children can be attributed to the increase in the land area. But how much? The answer, or at least some approximation of the answer, can be found in the base counts.

The Census 2000 retabulated total population based on the 2003-04 boundary for Wheatland J1 is 4,034, a difference of 1,331 people compared to the retabulation based on the 2001-02 boundaries. Therefore, a net 1,331 of the net 1,510 total population increase, 88.1 percent, can be attributed to geographic change. (See Table 4.)

Table 6. Comparisons of Income Year 1999 and 2002 Base Counts and Estimates for Wheatland J1 Elementary School District

1999 Income Year		**2002 Income Year**		**Differences**	
Total Base Population (Census 2000 population retabulated to 2001-02 boundaries)	**Total Population**	**Total Base Population** (Census 2000 population retabulated to 2003-04 boundaries)	**Estimated Total Population**	**Between 2001-02 base counts and 2003-04 base counts**	**Between income year 1999 estimate and income year 2002 estimate**
2,703	2,703	4,034	4,213	1,331	1,510

A similar analysis can be done for the change in estimated number of relevant poor children in families. Of the estimated 2,718 poor children in Kenosha County in Census

2000, four of them, or 0.15 percent, were geocoded to Wheatland J1 school district. When the same Census 2000 data were retabulated to the 2003-04 school district boundaries, 33 of the 2,718 total poor children in the county, or 1.2 percent, were tabulated within the Wheatland J1 district. The distribution based on the 2003-04 boundaries was used to produce the income year 2002 estimate for the district. The income year 2002 model-based estimate of poor children in the county is 2,800; approximately 1.2 percent of the 2,800, or 34 children, was estimated for the district. The base counts can be used to show how the boundary change altered the distribution of estimated poor children among the school districts within the county. Whereas Wheatland J1 was assigned 0.15 percent of the county's poor children based on the 2001-02 boundaries, the change reported in the 2003-04 boundaries led Wheatland J1's share to increase to 1.2 percent of the poor children within the county.

Again, it is important to remember that these data are all estimates, with some amount of error attached to them.[1] Nonetheless, the base counts do give data users at least some guidance as to how much effect geographic change is having on a population change in the area.

Conclusion

By understanding how geographic information is used in creating the estimates, data users will be better informed regarding how to appropriately use the data. Furthermore, if data users plan to compare data for the same areas over time, they should be aware of the impact of geographic changes to the areas, as well as other methodological issues documented by SAIPE. The Census Bureau made geographic updates to almost 25 percent of all school districts based on updates reported for the 2003-04 school year. In many, if not most, cases these changes ultimately had an impact on the total population estimates for the districts and possibly the estimates of the number of poor children in families. Retabulations of Census 2000 data to the updated geography can give data users a sense of the magnitude of these changes on the population, and aid them towards a better interpretation of the data.

Appendix A

The Geographic Update System to Support Intercensal Estimates

There are three main components that enter into GUSSIE: The Topologically Integrated Geographic Encoding and Referencing System (TIGER®), a database containing geographical information including address ranges; the Master Address File (MAF), which contains a complete list of all addresses and housing unit locations; and the decennial Census detail files, which contain the individual census records including addresses or location. Every unit on the MAF is given a MAF identification code. Those codes are also found on the decennial census detail files, allowing the files to be matched.

Boundary changes at any level of geography are reported to the Census Bureau and recorded in TIGER®. TIGER® is linked to the MAF, and census block and other geocodes in the MAF are updated to reflect the changes in TIGER®. The updated MAF is then matched to the Census detail files based on the MAF identification code. A new version of the detail file is created containing current geocodes, and the census is thus retabulated.