

Your State Data... Your Local Story

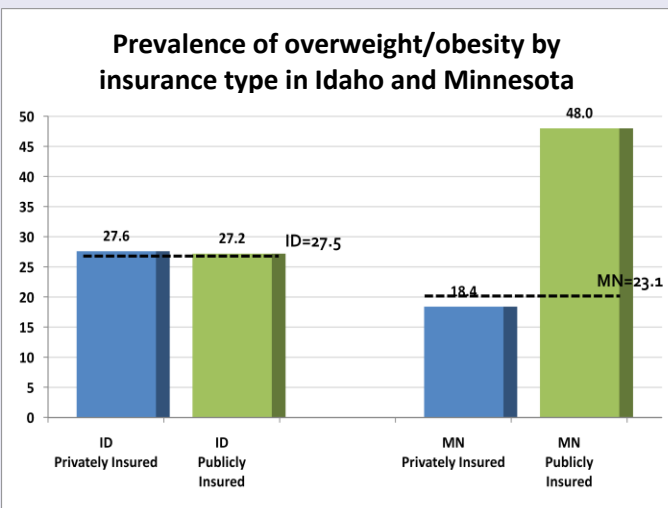
Local Uses of National and State Data

And how to construct a synthetic estimate

Do you always need local data?

No! In fact, national and state data can often be applied locally and have many local uses:

- **Reforms needed at the state level are likely also needed at the local level** – this isn't likely to change with slight prevalence differences
- Combined with what is already known about your local area, **state level data can be very powerful** in informing change and measuring benchmarks
- **Data collection is expensive** – consider what you can do with the data and information already available
- Local data make up state estimates. If demographic distributions between a local area and the state are similar, state and local estimates likely are too. However, large within-state demographic variation may mean that local areas actually differ markedly from the state as a whole. In these cases, a **synthetic estimate** can help provide a more accurate local picture.



However, large within-state demographic variation may mean that local areas actually differ markedly from the state as a whole. In these cases, a **synthetic estimate** can help provide a more accurate local picture.

The graph to the left is an example of when summary measures do not tell the whole story. In Idaho, the state overweight/obesity prevalence is quite similar to that for both privately and publicly insured children within the state. However, in Minnesota that is not the case. While Minnesota has a lower overall prevalence, it has much greater disparities in overweight/obesity by insurance type. We would not have known this had we not stratified by an important subgroup.

Similarly, local areas within a state can vary on factors known or suspected to affect health, health care and the other topics in the NSCH and the NS-CSHCN. Synthetic estimates can help us understand the details of what is happening at the local level and thus inform policy, program reform, and improvement efforts without the expense of collecting new data.

What is a synthetic estimate?

A synthetic estimate is a prevalence estimate for a local area that is calculated by using descriptive or demographic data for local areas combined with state prevalence values. It is similar in concept to an indirect adjustment. A local estimate is most likely to differ from a state estimate if the demographic distribution at the local area differs from that of the state AND the prevalence of the indicator varies by the same demographic factor.

For example, medical home varies by race (1st column) and the race distribution in Marin County differs from California as a whole (2nd and 3rd columns). 49.6% of all children in California are estimated to have a medical home, but it's possible that the estimate for Marin County is different.

	% with a medical home by race in CA	Race distribution in California	Race distribution in Marin County
Latino/Hispanic	37.6%	49.3%	34.5%
White	65.7%	30.6%	54.9%
Black	42.2%	5.8%	2.4%
Multiracial	71.0%	3.7%	4.6%
Other	50.6%	10.7%	3.6%
All Races	49.6%	-	-

Data Sources: childhealthdata.org and kidsdata.org

So, let's calculate a synthetic estimate!

We'll estimate the percentage of children in Marin County with a medical home.

STEP 1: Determine the prevalence of your variable by selected demographic category at the state level. You can choose any variable for which you have state-level data.

www.childhealthdata.org provides data on numerous measures of child health and well-being and allows stratification by various subgroups. We used data from the 2007 NSCH to find the prevalence of having a medical home in California stratified by race/ethnicity.



STEP 2: Determine the number of children in your county who fall into each category of the demographic characteristic you are using. You can use any demographic variable for which you have county and state-level information.

Race/Ethnicity Category	Distribution in Marin County
Latino/Hispanic	16,241
White	31,583
Black	1,269
Multiracial	2,570
Other	1,968
Total	53,631

We got the 2007 race distribution in Marin County directly from KidsData.org (California only).

Note that we combined the Native American and Asian/Pacific Islander groups from the KidsData website into an "other" category to match categories in the 2007 NSCH. *It is important to make sure the groupings in your two data sources match!* You can also access county-level information from places such as: www.KidsCount.org, www.census.gov and your state department of finance.

STEP 3: Calculate the estimate. First, determine the estimated number of children who meet the indicator of interest within each demographic group for your selected county. In this example, it is the number of children with a medical home by race in Marin County (3rd column in the table below).

Then, determine the prevalence of your variable of interest in your county by dividing the total number of children in the county who meet that variable by the total number of children in the county. Here, we divide the total number of children estimated to have a medical home in Marin County by the total number of children living in Marin County in 2007: **30,214/53,631=56.3%**

Race/Ethnicity Category	Distribution in Marin County	% with medical home by race in CA	# with medical home by race in Marin County
Latino/Hispanic	16,241	37.6%	16,241*0.376= 6,107
White	31,583	65.7%	20,750
Black	1,269	42.2%	536
Multiracial	2,570	71.0%	1,825
Other	1,968	50.6%	996
Total	53,631		30,214

The prevalence of children with a medical home in Marin County (56.3%) is estimated to be higher than the prevalence in California (49.6%). This makes sense intuitively because the race/ethnicity category that has the most children with a medical home is more prominent in Marin County than in California as a whole.

Be aware of bias and limitations of synthetic estimates.

- Timing differences between when state and local data were collected
- Differences in the way demographic categories were defined in each data source - what is labeled the same may not actually be the same (race is an especially subjective category)
- Synthetic estimates done this way are based upon one demographic factor and don't account for variations by other demographic categories that may also be important in explaining your variable of interest
- Number of stratifying groups is limited by state sample size
- State data with high relative standard errors cannot provide reliable synthetic estimates
- You cannot construct confidence intervals nor do statistical tests with synthetic estimates