



## Questions and Responses

*regarding review of the 2017 population projections methodology*

### **Q1: How does methodology used for the 2017 projections differ from the one used for the 2012 projections? In what ways does the 2017 methodology improve on the prior approach?**

The 2012 and 2017 methodologies are completely different, selected because of different assumptions about what projections were most needed (near-term v. long term) as well as what input data was available. The 2012 methodology involved sophisticated modeling, driven by statistical theories. This approach was selected to focus on the 30-year trend into 2040 (and therefore based on observations from the past several decades). The 2017 methodology, on the other hand, focuses on near-term projections to 2020, and uses an extrapolation method based on the most recent Census and estimates.

The work for the 2012 projections started in 2011, when the 2010 Census results became available. As it was the Cooper Center's first opportunity to prepare the statewide projections for 10, 20, and 30 years forward, we opted to focus on making the 30-year projections as reasonable as possible. According to projections theory and literature, long-term projections require input data that reaches back as far into the past, if not longer, than the projections will reach into the future. We researched extensively various methodologies and selected a highly sophisticated statistical model that took a long view across time and localities over several decades, and identified patterns among all Virginia localities, as well as within each individual locality. It was an organic, big stroke (deep past, long future) statistical modeling approach. The focus was the future in 30 years.

We have since learned from data users that: 1) localities tend to use projections most frequently when looking at the near, instead of the distant, future; 2) an intuitive, easy-to-understand methodology is preferred to those involving advanced statistics approaches; and, 3) assumptions and data based on the recent past are preferable to those based on years longer past.

Midway through the decade, with available estimates indicating trends since the 2010 Census, we have researched thoroughly and developed a new methodology in which we attempted to be responsive to the needs of data users. Of course, our ability to respond to specific needs of individual localities is limited by ***the guiding principle underlying these projections, which is to use a uniform methodology and consistent input data to produce projections with small mean absolute percentage error (MAPE) across all localities.***

Compared to the 2012 methodology, the 2017 methodology is straightforward, and can be easily implemented or replicated by data users. The assumptions are:

1. For the 2020 projections, we calculate the annual growth rate between 2010 and 2015 for each individual county/city. We assume this growth rate will continue for the rest of the decade (through 2020).
2. For the 2030 and 2040 projections, we calculate average annual growth (the absolute number, not the rate) between 2000 and 2015 (a 15-year span) and assume that amount of growth will continue every year throughout the next two decades, beginning in 2020 and 2030.

As a final step in the process, all locality projections are proportionally adjusted to add up to the state projected total.

We believe the 2020 projections using the 2017 methodology are likely to be more accurate than the projections resulting from the 2012 methodology primarily because more recent input data are available. For projections farther into the future (such as 2030 and 2040), it is harder to know which approach produced smaller MAPE across the entire state. We've certainly seen cases where population trends reverse, especially over a longer time period, so "accuracy" - in terms of predicting what happens in the future - is likely to diminish over the course of next 25 years. But that is true of all projections and all methodologies. We believe the 2017 methodology is an improvement over the 2012 methodology for data users experience, because it does not incorporate potential reverse trends, nor does it take into consideration the speed of acceleration or deceleration typically observed over a longer time period.

The real accuracy of projections can only be measured when we have the 2020 Census data. Our goal is to update projections on a more regular basis, with less time between rounds. This will allow the projections to reflect the most recent trends, which is critical to improving accuracy.

## **Q2: We would prefer the Cooper Center to use our locality-produced 2015 estimate instead of Virginia's official population estimates produced by the Cooper Center. Can that be done?**

In order to meet the requirements for comparable statewide projections, what we produce must be developed using a consistent methodology, and only input data available in all localities. Locality-produced estimates may be derived using different methodologies and different input data. While these locality-produced estimates may be valid for local planning purposes, they are not the official estimates for the state. An alternative to using the official state population estimates, produced by the Cooper Center, would be to use the official Census Bureau's annual population estimates. If a majority of the localities would prefer that we use the Census estimates, we could consider switching. Given the stipulations established for these official projections, our professional standards do not allow us to select different sources of estimates data for different localities.

## **Q3: How do you validate the 2015 Cooper Center estimates which are used in the 2017 projections as input data, compared to estimates produced by other entities?**

Different methodologies and input data produce different estimates, and they can all be valid. Since estimates are not Census headcounts, they are always either over or under the true numbers. The only way to determine which set of estimates is most accurate is to compare the 2020 estimates to the 2020 Census. It is worth noting that validation of statewide estimates is not so much about each individual locality as about the average error across all localities.

## **Q4: How does the Cooper Center develop its population estimates?**

The estimates methodology is based on a commonly used method called "ratio-correlation". The input data include births, deaths, school enrollments, housing stock, and driver's licenses. A detailed methodology document is available on our website at <http://www.coopercenter.org/demographics/virginia-population-estimates>.

## **Q5: The Census Bureau has a Challenge Program for its estimates. Would the Cooper Center establish something like this for projections?**

Yes, there is a review/challenge period following the release of provisional population projections. Our appeal process for projections (and for estimates) is precisely modeled after the Census Bureau's Challenge Program

for population estimates, which, on the Bureau’s website states: “A challenge may result in a revised estimate if either of the following errors are found:

1. Technical errors in processing input data or producing the estimates
2. Incorrect input data used in the process of generating the estimates”.

This approach does not accept as challenges the results of alternative methodologies or of alternative input data. It does not necessarily mean the alternatives are not valid; it only means that using alternatives does not meet the principle of using the same methodology and same input data for all.

### **Q6: How would you handle outliers, results which seem impossible?**

A wide range of projected outcomes—based on different input data, methodologies, and assumptions—are highly possible and can all be valid, as the future is inherently unknown. Outliers are typically defined as unusual and implausible outcomes. The methodology for the 2017 projections is designed in an elegant way to eliminate outliers, as all projections are extrapolated based on tangible numbers observed in the near past.

### **Q7: Would you be able to provide a high-mid-low series of projections to accommodate needs for a range?**

If future contracts allowed additional funding to project high and low series, we would be able to expand the projections into a range. Again, as guiding principle, the methodology and input data for producing the high and low range would need to be uniform and consistent across all 133 localities.

### **Q8: How do you validate the projections and measure errors?**

We plan to conduct a thorough evaluation once the 2020 Census results are available. Census numbers remain the only gold standard, as they are actual counts.

It may be helpful to know that, in the development phase of this process, we tested the methodology by applying multiple sets of prior Census and estimate data, and evaluated the projected numbers against the actual Census numbers. As one example, we used the 1990 and 2000 Censuses and the 2005 Cooper Center estimates in the current methodology to project 2010. We then compared the resulting 2010 projections to the 2010 Census counts. The MAPE across all localities was 4.41. This compares highly favorably to what would be expected as a standard MAPE of 12 percent for county level 10-year projections.

### **Q9: Are projections used by state agencies for allocating funds to localities?**

We have not found any official document or evidence to suggest that projections are used by state agencies for allocating funds. In the *Code of Virginia*, the words *estimates* (for the current or past) and *projections* (for the future) are sometimes used interchangeably, which results in confusion. For example, some references to Weldon Cooper Center projections actually are meant to be Cooper Center estimates. This is evident when you realize that all projections before 2012 were produced by the Virginia Employment Commission, or its contractor.

### **Q10: Could you conduct a statewide users survey on how the projections are used?**

We would be interested in conducting such a survey to help better understand how projections are used by a wide range of users. The survey, a costly undertaking beyond our current scope of services, would require a substantial amount of funding to support the efforts of designing a scientific questionnaire, identifying users with representation, following up with non-responses, and analyzing and reporting results.