

## **Summary of Testimony on Census Undercount for the House Subcommittee on the Census**

Kenneth J. Darga, Senior Demographer  
Michigan Department of Management and Budget  
May 5, 1998

I would like to thank Chairman Miller and all the members of the Subcommittee on the Census for inviting me to speak with you today about Census undercount adjustment. At this time I would like to submit two papers for the record which I will summarize briefly.

### **The Fallacy of Undercount Adjustment**

It's no surprise that the Census doesn't count everybody. The Census has a hard time counting people who don't trust the government or don't want the government to know where they are. The Census doesn't do a very good job counting homeless people either, and there are many other factors that make a complete count very difficult.

So the Census Bureau tries to fix the problem by counting people in some neighborhoods a second time and comparing the results person-by-person with the Census. *In 1990, this method seemed to find just about all the people in this sample of neighborhoods who were missed by the Census.*

This sounds great until you realize what's really happening. The 1990 Post-Enumeration Survey didn't really find all the people that were missed. People who didn't want to be counted the first time didn't want to be counted the second time either, and the Post-Enumeration Survey didn't even try to count homeless people. But it did find quite a few people who looked like they were missed by the Census when they really weren't. In fact, *most* of the people that the Post-Enumeration Survey identified as missed by the Census really weren't missed by the Census.

That's a surprising claim. How can you know that it's true?

There are at least two ways: a theoretical approach, and an empirical approach.

### **Theoretical Verification**

First, a theoretical approach. On pages 6 through 9 of my first paper,\* you will find a very simple and very basic statistical phenomenon that explains why serious problems are inevitable when you try to measure undercount with a coverage survey. These pages show that an effort to measure a small component of the population--such as people missed by the Census--is very sensitive even to extremely small sources of measurement error, and that the coverage survey has to contend with a lot of very large sources of measurement error.

\* Kenneth J. Darga, "Straining Out Gnats and Swallowing Camels: The Perils of Adjusting for Census Undercount." Submitted to the Subcommittee on the Census, House Committee on Government Reform and Oversight, May 5, 1998.

So it shouldn't be surprising that the coverage survey identifies a lot of people as missed by the Census when they really weren't. It would be a lot more surprising--unbelievable, in fact--if it didn't.

## **Empirical Verification**

You can also see the problems with the undercount adjustments by taking an empirical approach. The Census Bureau evaluated the 1990 PES quite extensively, and it did a very impressive job of documenting its shortcomings. I also want to acknowledge the important work of Dr. Leo Breiman of the University of California at Berkeley in evaluating the Census Bureau's evaluations.

My second paper\*\* discusses six very serious sources of error that were documented by the Census Bureau:

- survey matching error
- fabrication of interviews
- ambiguity or mis-reporting of usual residence
- geocoding errors
- unreliable interviews
- unresolvable cases.

And the Census Bureau didn't document just a little bit of error. One thing that the theoretical approach and the empirical approach have in common is that they both demonstrate very large amounts of error in the Census Bureau's adjustments for undercount. *The adjustments based on the Post-Enumeration Survey reflect errors in measuring undercount even more than they reflect undercount itself.*

## **Impact on Census Data**

Now you might think that, since the estimated net undercount is less than two percent of the population, even a bad adjustment for it wouldn't cause big problems. Before you make that mistake, it is important to consider the examples on pages 11-15 of my first paper.\* These pages demonstrate that the 1990 PES identified some undercount differentials of 10 percentage points, 20 percentage points, and more that turned out to be totally spurious.

I want to be clear about what I mean by a difference of 20 percentage points. These examples don't just involve inflating one group by 1% and another by 1.2%. That would be a difference of 20 percent. If the difference should really be 0 percent, that could be a problem for some purposes. But that is not what I mean by a difference of 20 percentage points.

\* Kenneth J. Darga, "Straining Out Gnats and Swallowing Camels: The Perils of Adjusting for Census Undercount." Submitted to the Subcommittee on the Census, House Committee on Government Reform and Oversight, May 5, 1998.

\*\* Kenneth J. Darga, "Quantifying Measurement Error and Bias in the 1990 Undercount Estimates." Submitted to the Subcommittee on the Census, House Committee on Government Reform and Oversight, May 5, 1998.

These examples involve inflating one population group by 8% and another group by 28% when neither group has been undercounted more than the other. This is not a problem that only demographers would be concerned about: This problem is big enough to affect every user of Census data.

In an effort to solve an undercount of less than 2%, the reliability of the Census would be utterly destroyed. This is a strong statement, but that does not mean it is an overstatement. It would be very difficult to overstate the implications of having errors of this magnitude integrated with the Census counts.

Thank you for the opportunity to testify this afternoon. I would be happy to answer any questions you may have.