

Discussion of Census Coverage Measurement Session Lynne Stokes, Southern Methodist University

These four papers are of two distinct types. The first three were about the operational details of the coverage measurement program, and the fourth one about an experiment and its analysis. Therefore, the types of comments I have about the two groups of papers are quite different.

I'll start my remarks with the first three papers. I am impressed with the plans that led to such rich data being collected from the coverage measurement program. While the purpose of these papers was to explain how that was accomplished, I felt like I got to the middle of a novel and then didn't get to finish it. I'd like to see more analysis and interpretation of the data, so most of my comments will be about that.

The purpose of the coverage measurement program was explained this way:

“The purpose of the 2010 CCM program was to evaluate coverage error in the 2010 Census to determine what needs improvements in future censuses ...”

I take that to mean improvements in both enumerations and coverage evaluations in future censuses, so I have some comments about both.

I will start with the first paper, “Results of the 2010 Census coverage Measurement Field and Matching Operations,” by Sanchez, Wakim, and Cronkite. This paper contained a very helpful overview of the many operations that were a part of the CCM, and how they fit together. When I first looked at these papers, I didn't begin with this one, and got very confused about which operation was which.

My first comment really applies to all of these operations papers, but I'll mention it here because this one is first. In many cases, I wanted the answer to this question: What did a particular operation accomplish? The table below is from one of the papers and is a typical presentation.

Table 1 Match Codes Following Each Stage of Matching (Percent of Total)

	P-Sample : 171,217 HU's		E-Sample: 180,528 HU's	
	Before Followup	After Followup	Before Followup	After Followup
Matches	94.78	95.04	88.17	88.40
Possible Matches	0.07	0.01	0.05	<0.01
Nonmatches	2.59	2.47	6.32	5.38
Duplicates	0.01	0.01	1.33	1.59
Not a HU	2.54	2.47	4.13	4.62

Let's focus on the P-sample housing units. If you compare the before field follow-up and after field follow up data, it looks like the follow-up didn't do much, since the percentages in each category did not change much. But this may not be true, and there is no way to tell from this presentation. Here are two tables, either of which could be true,

since they have the same marginal as the data that was presented in the previous table. But each of the two of them tells a different story about what the field follow-up accomplished. If the first table is true, it says you might as well have not done the follow-up, while the second table says that the follow-up made more changes. Which is the case is informative about the necessity of the operation, while the marginal themselves are not so helpful.

Alternative table 1: Presentation of follow-up results showing little change

		AFU					
		Match	Possible match	Non-match	Dup	Not HU	Total
B F U	Match	94.78	0	0	0	0	94.78
	Possible match	0.06	0.01	0	0	0	0.07
	Non-match	0.09	0	2.47	0	0.03	2.59
	Dup	0	0	0	0.01	0	0.01
	Not HU	0.11	0	0	0	2.47	2.54
Total		95.04	0.01	2.47	0.01	2.47	100

Alternative table 2: Presentation of follow-up results showing substantial change

		AFU					
		Match	Possible match	Non-match	Dup	Not HU	Total
B F U	Match	90.08	0.01	2.34	0.01	2.34	94.78
	Possible match	0.07	0	0	0	0	0.07
	Non-match	2.46	0	0.06	0	0.06	2.59
	Dup	0.01	0	0	0	0	0.01
	Not HU	2.41	0	0.06	0	0.06	2.54
Total%		95.04	0.01	2.47	0.01	2.47	100

Another way to think about the relative value of each operation would be to try to put their value all on the same scale somehow. For example you might try quantifying (predicting) how the final estimates of the components of coverage error would differ if each operation were eliminated. Or perhaps it would be more sensible to scale this value by how much that operation costs, to get a measure of efficiency of each operation.

Now I will turn to the Cantu and Johnson paper. First, I was amazed by the thoroughness of the search for duplicates this time. I appreciated the example of Billy Bob in the paper because he helped me process all the possibilities a lot faster. The two comments I made about the first paper apply here too—about data presentation and evaluating the value of each operation in the search. But now I will turn to some additional questions.

There was little said in the paper about how the findings will help improve the next Census. One way that could have been done was to include more description of who these duplicated people were, and what that suggests about how the census should be operated. For example, suppose the reasons for duplicates found were distributed as shown in Figure 1a. Then this might suggest that adding a specific question or at least instructions about kids in dual custody to the census form would be reasonable and cost-effective. On the other hand, if the data looked more like Figure 1b, where several types of people were equally likely to be duplicated, it might suggest that you couldn't really afford that much real estate on the census form. So basically, more description of how these results inform the census operation would be interesting and valuable.

On the other hand, if the data looked like Figure 2, where several types of people were equally likely to be duplicated, it might suggest that you couldn't really afford that much real estate on the census form. So basically, more description of how these results inform the census operation would be interesting and valuable.

Another idea for improving the efficiency of the evaluation process might be to build a model that predicts the likelihood that two possible matches are a match or nonmatch, and then use that to decide when to send a case out to the field. If it is very likely to be a match or nonmatch, then a visit isn't necessary, but otherwise it would be. Alternatively, you could use the model to improve estimation of duplicates in the census in the non CCM blocks, by using the reciprocal of the estimated probability as a weight. There are many features of a possible match that might be informative, such as unusualness of the demographic characteristics, proximity of the addresses, or stated relationships between the potential duplicate and other members of the household.

Now I will move to the final operations paper. This paper did include more detail about some of its findings that could be informative about how to improve census operations. For example, we learned that the housing units missed were the moveable ones, like tents and mobile homes not in a park. And we heard that asking probe questions could turn up the very people we are at risk of missing, the young and minorities. But still, it was hard to draw a direct line between what was reported and how exactly to improve the census. For example, would it be worthwhile for the census canvassers to interview the household to find those hidden housing units? To answer this, you would need to know how much it costs and also how much difference it makes. For example how many of the people in those added housing units were counted anyway. If people living in tents in the

backyard are usually reported as being in the household anyway, then maybe getting the housing units right is not so important. Alternatively, maybe the CCM data could be used to determine a profile of the address types that do have hidden units, so that the canvasser could be required to interview only those addresses.

As another example, on the household interview, the data might be informative about which of the annoying probe questions should be retained. This would depend on the improvement in coverage each provides, which would require knowing how many extra people each found and how many duplicate people each found. Though duplicate people are not as bad as missed people, I would think you would want as few as possible for efficiency reasons.

Finally, I turn to the last paper about the recall experiment. First, I thought the use of change of address records for coming up with a reasonable sample was a very clever idea. It occurs to me that this might be a good source of alternative addresses in the search for duplicates too. This might have already been done, but if not, maybe it would be worth the effort to try to assess how effective this would have been for the CCM sample.

I have a few comments and questions about this paper. First, when I look at the table below, which is from the paper, I don't think the interesting thing is the change or lack of change in the proportion reporting a move, but rather, why do so few report a move at all? This seems like something that at least deserves a comment, and at most might help explain why people can't report moves accurately.

	Panel 2 (June 2010)	Panel 3 (September 2010)	Panel 4 (February 2011)	Overall
Mover	34.41	32.64	31.54	32.98
Long-term Cyclers	15.75	16.60	8.49	13.94
Unknown	2.23	2.09	2.91	2.38
Non-Mover	47.61	48.67	57.06	50.70
Data Source: RBS Output				

Another detail that was little discussed in the paper was the success of the address and phone matching operation was. In my experience, many of these addresses are unmatchable, due in part at least to increasing cell phone use. I would expect this to be particularly problematic among recent movers. So the bias that could occur due to this source of nonsampling error might confound the comparison over time.

My next comments are about the use of a dual frame design where the two frames were the landline and cell phone. It is becoming increasingly important to include cell phones since there are so many non-landline households, and I suspect that this is especially true for people who move often. However, I don't think weighting is really necessary in this case. You aren't trying to estimate a nationwide incidence of remembering a move, but rather are just trying to compare the proportion that remember under different conditions.

I think just comparing the change over time within each population would be adequate. In fact the comparison would be cleaner. First, you wouldn't lose any of the sample due to item nonresponse, as you do due to the telephone ownership question. Secondly, the differential response rate over time allows for the response bias to vary, making it hard to determine what is a real change in recall rates and what is due to a change in the response bias differential.

Finally, I enjoyed learning about the CCM program. There is much information about improving the next Census hidden in the data collected, but it will have to be dug out. I hope that the funding exists to continue the search.