DISCUSSION

Tommy Wright, Oak Ridge National Laboratory
ORNL, P. O. Box 2008 Bldg 6012, Oak Ridge, TN 37831-6367

It is indeed an honor for me to have been invited to speak briefly on these four excellent and information packed papers. I agree with much that they contain. Since 1980, I have been intrigued by the fact that the most fundamental mathematical operation of *counting* can be so difficult to implement when it comes to correctly enumerating a finite set of persons.

These papers ask us to look ahead at the **1995 Test** plans to be implemented at four sites. The Bureau, as reflected by these papers, is to be commended for its good *scientific practice* of continuing to use an *OPEN PROCESS* in developing and carrying out these plans. Often our vision for the future, 1995 and eventually the 2000 Census and beyond, can be enriched if we periodically look to the past. To facilitate a quick look back at 1990, I have prepared a *QUIZ* for the audience.

POP QUIZ

I. What has been reported and accepted as the official resident population of the United States as of April 1, 1990?

(a) 253,393,786 (c) 253,979,141
(b) 248,709,873 (d) 252,712,821

II. What is the believed difference in the net undercounting between African-Americans and non African-Americans in the reported 1990 U. S. Census?

(a) 0 % (b) about 1/2 % (c) 3 to 4 %

III. What is the significance of the number $2,600,000,000.00 ?

(a) The approximate cost of the 1990 U. S. Census.
(b) The amount that an individual would have earned at a rate of $2.52/minute (without interest) if he/she had been counting nonstop with check marks since the reported death of Jesus Christ.

(*KEY*: 1. (b); 2. (c); 3. both. Audience's Score: 100%.)

This pop quiz is instructive.

Lesson 1: Provide Only One Number

Reporting more than one number for the census will always lead to controversy. With more than one number *and* scientific evidence that one of the numbers, other than the direct count, is superior to all others, the methodology used to produce this superior number should be *integrated* with the direct counting process to lead to one best number. This is the Bureau's current thinking behind INTEGRATED COVERAGE MEASUREMENT(ICM) leading to a *ONE NUMBER CENSUS* with which I agree.

Lesson 2: Decrease All Undercounting

A controlled mass enumeration along with an extra high quality sample enumeration which makes use of *statistically defensible* sampling and estimation methods, as would be the case with either **CensusPES** or **CensusPlus**, is a good strategy to achieve greater accuracy. When a sample of areas is enumerated on more than one occasion and these separate enumerations are carefully reviewed against each other and possibly other evidence to yield a single *"resolved"* enumeration for each of these sample areas, the final resolved counts for these sample areas are improved. If this improvement can be *shared* with the nonsample areas, as under the two ICM variations compared below, there will be overall improvement in the enumerations.

ICM Variation I-CensusPES: Assume two surveys (Mass Enumeration & PES Sample Enumeration) are made of a population with M blocks at about the same time. With independence of the two surveys, the **unknown** total population size N can be estimated by

$$\hat{N}_{CPES} \equiv \frac{\begin{pmatrix} Mass \\ Count \end{pmatrix}\begin{pmatrix} Estimated\ POP\ Count \\ Based\ on\ PES \end{pmatrix}}{\begin{pmatrix} Estimated\ Count \\ In\ Both \end{pmatrix}}$$

CensusPES is aiming in the right direction, but independence violations, timing, matching, and the assignment of nationally estimated undercount rates down to smaller areas remain points of concern (Hogan, 1992; Mulry and Spencer, 1991). In an effort to overcome these concerns, CensusPlus is being considered.

ICM Variation II-CensusPlus: Assume two surveys (Mass Enumeration & "Plus" Sample Enumeration) are made of a population with M blocks at about the same time. The plus sample blocks undergo a second *extra high* quality count which when compared with the initial count leads to **observed** resolved counts for the sample blocks. Under a simple model (Wright,1993), resolved counts are **predicted** for the nonsample blocks. Hence an optimal estimator of N is obtained by adding these observed (in sample) and predicted (not in sample) resolved block counts. In fact, this sum turns out to be the *classical ratio estimator*

$$\hat{N}_{CPlus} \equiv \frac{\begin{pmatrix} Mass \\ Count \end{pmatrix}\begin{pmatrix} Estimated\ POP\ Count \\ Based\ on\ PLUS \end{pmatrix}}{\begin{pmatrix} Estimated\ POP\ Count\ Based\ on \\ Mass\ Counts\ of\ PLUS\ Sample \end{pmatrix}}$$

Thus, the two ICM variations, being tested in 1995, have similar looking estimators which compensate for net undercounting in the mass enumeration. For more on $\hat{N}_{CPlus}$, some of its potential advantages, and some early encouraging empirical results, see Wright(1993,1994).

Simplicity and *universal understanding* of whatever methodology is to be used by the statistical scientists and political scientists **before the act** is absolutely critical and essential.

Lesson 3: Control Costs

A top goal of the Bureau is to be frugal in its improved counting. Having to follow up nonresponding households is a big census cost. Why? Have we failed to effectively convey the importance to the country and to the individual of this decennial data collection effort? Do the census procedures discourage timely responses? Are the census questionnaires easy to complete and sent to the right addresses? Do we attempt to capture too much information from too many? The following five innovations which address these questions have potential for controlling 2000 Census costs.

1. Create and Maintain a Master Address File with Periodic U. S. Postal Service Updates (paper by Gbur)

All efforts should be devoted toward keeping the addresses straight, for this listing is the foundation for a successful census. Many benefits would be realized if, as planned, there are local government reviews of *actual* addresses before mailing questionnaires.

2. Improve Method for Listing Household Members and Simplify Census Questionnaire (paper by Martin)

Martin makes it clear that getting a list of household members with census questionnaires is not easy - especially for complex living arrangements. Use of *probing* questions to build up a household listing shows promise; subsequent efficient trimming remains a challenge. This important research should continue.

3. Deliver the Questionnaire Differently (papers by Thompson/Mulry)

Bravo to efforts by Don Dillman and colleagues for the sequence: (i) a prenotice letter, (ii) the initial questionnaire, (iii) a reminder postcard, and (iv) a thank-you *or* a replacement questionnaire. This sequence of events should occur over a brief period of time to convey *URGENCY* to the public. Is there any hope for three weeks instead of six?

4. Focus on Short Form Data; More Data Later with CONTINUOUS MEASUREMENT

5. Follow Up Only a Sample of the Nonrespondents(papers by Mulry/Thompson)

If significant nonresponse persists, the Bureau is testing two sampling approaches based on research by Cary Isaki and colleagues which estimate the nonrespondents. In the short term, this fundamental change may save money; but the long term costs, not always monetary, remain unclear and require study. While having 40% mailback, followed by a 1/3 sample of nonrespondents yielding an additional 20% response and estimating for the remaining 40% may be acceptable to a statistical scientist, its acceptability to a political scientist is in doubt. I am a firm believer in F. F. Stephan's 1948 statement that, "...all scientific observation, whether statistical or not, is based on sampling." However, we should pause to recall the teachings of Leslie Kish(1983). From Table 1, we see that our losses in details, coverage, completeness, and credibility are positively correlated with the amount of nonsample nonresponse follow-up. Professor Kish also notes, "A cautious statistician does not undertake lightly the task of comparing samples with censuses, and, especially, he (or she) does not suggest lightly the possibility of substituting one for the other. There are heavy stakes involved with each census - - financial, legal, constitutional, professional, and emotional stakes." We must heed these words, be cautious, and think more deeply about the consequences of sampling for nonresponse follow-up.

Table 1. Comparisons of SAMPLES and CENSUSES

Criteria	Samples	Censuses
- Rich, Complex, Diverse, Flexible	* * *	
- Accurate, Relevant, Pertinent	*	
- Inexpensive	*	
- Timely, Opportune, Seasonal	* *	
- Precise(large and complete)		*
- Inclusive(coverage), Credible, PR		*
- Detailed		* *
- Population Content	* *	*

*Denotes Advantage. Source: Kish(1983).

RECOMMENDATIONS

I. Speed Up The Process; Convey Urgency

I believe that most if not all nonresponse is caused by inadequacies in the data collection system, not the nonrespondents. I also believe that earlier returned questionnaires tend to have better quality data than later ones. Thus we are more likely to meet Judith Lessler's call for an unending pursuit of 100% response if we shorten the *field* data collection time.

Recommendation: The Bureau should commit to finding and executing sincere efforts which will convey urgency and motivate every person in the United States to have responded by April 1, 2000.

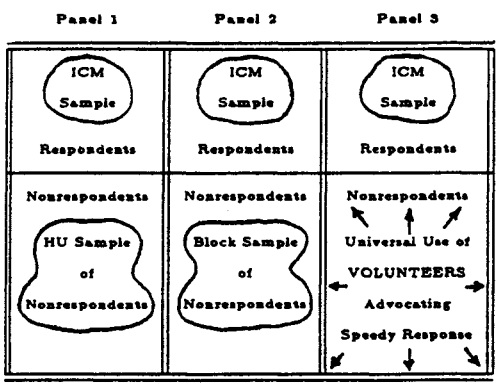
II. Consider Using Motivated VOLUNTEERS Instead Of or in Conjunction with Sampling for Nonresponse Follow-up

The ICM sample seeks to decrease differential undercounting and improve overall coverage. The nonresponse followup sample seeks to hold down immediate monetary costs with an increase in uncertainty and a loss of quality in detail, coverage, and completeness. How do we establish the point at which sampling for nonresponse follow-up begins? If we wait until a small

level of nonresponse remains, do we need both the ICM sample and the nonresponse followup sample? I believe that there will be a need for some form of universal follow-up of nonresponse beyond the reminder postcard and the replacement questionnaire.

Recommendation: One alternative to sampling for nonresponse follow-up for saving money is to consider testing the use of widely scattered motivated VOLUNTEERS. Being from Tennessee and having recently been a volunteer when my town of Farragut (approx. 15,000 persons) conducted a 1994 census with volunteers going door to door in their own communities, how can I avoid the natural thought of volunteers? Because I was assigned to enumerate 18 homes (51 persons) on two streets near my home, my neighbors knew my face; I knew their faces; and I knew when they were at home. In a total of less than 3 hours stretched over a one week period, I achieved 100% responses for every one of my assigned addresses. The Bureau has heard over and over again from local and community people to make more use of them. One way to test the use of volunteers in the 1995 Test is to add a Panel 3 to the currently planned 2 panels at each of the four test sites. Reconsider organizing volunteers for each

ALTERNATE 1995 Test Layout at Each Site



block (e.g. block captains / block facilitators as suggested in the Bureau's Toolkit) who would be contacted by the Bureau periodically throughout the *brief but thorough* followup period (not to exceed say two or three weeks) to personally remind and to encourage those nonresponding neighbors to respond. *These volunteers would not collect any data, but they might help replace questionnaires.* Can you imagine your Representative to Congress as a volunteer going door to door in his or her district as an *advocate* for the census?

III. Prune the Process

Recommendation: By pruning, our temporary loss results in an ultimate gain. The ONE NUMBER CENSUS, with its master address file, simplified and improved short questionnaire, concentrated efforts emphasizing the urgency of responding to the questionnaire, and integrated coverage measurement seem right on target. But the Bureau must continue to take a fresh objective view to further trim the process of operations that have a history or potential for inviting problems.

IV. Establish a Research Group which Is Removed from Day-To-Day Planning and Development with Possible Title - "Sampling and Estimation Methods for Census Taking"

The current efforts for the 1995 Census Test are highly focused towards development and getting a specific job done with less emphasis on comprehensive research, i.e., they are important *frontline* efforts. However, more careful thought and research are needed to support these frontline efforts.

Recommendation: With a contemplated more prominent role for *sampling* in the census process in various forms such as: (i) matrix sampling, (ii) sampling of nonresponding housing units, (iii) sampling for integrated coverage measurement, and (iv) sampling for continuous measurement, and because all is not known about sampling and estimation theory for finite populations, it is recommended that the Bureau consider the establishment of a Core Sampling Research Group. Such a research program is not a luxury, and it might: (i) define problem areas and pursue research in sampling and estimation methods associated with census taking, (ii) interact with and support the frontline efforts, (iii) carry on a dialogue with other research groups on planned census changes, (iv) be a source of collaborators for specifics relating to the frontline efforts, (v) help to develop an effective program for the education of the public if these changes are to be implemented, and (vi) involve graduate students in the research.

Many thanks to the authors; I enjoyed their papers.

References

-Hogan, H. (1992). "The 1990 Post-Enumeration Survey: An Overview," *The American Statistician, 46*, 261-69.

-Kish, L. (1983). "Data Collection for Details over Space and Time", in *Statistical Methods and the Improvement of Data Quality*, (T. Wright, ed), New York: Academic Press, 73-84.

-Mulry, M. and Spencer, B. (1991). "Total Error in PES Estimates of Population", *Journal of the American Statistical Association, 86*, 855-61.

-Wright, T. (1993). "CensusPlus: A Sampling & Prediction Approach for the 2000 Census of the United States", unpublished paper, U. S. Bureau of the Census, Washington, D. C.

-Wright, T. (1994). "CensusPlus: A Sampling & Prediction Approach for the 2000 Census of the United States, First Addendum(Update, Empirical Evidence, & Sample Size Determinations by State)", unpublished paper, U. S. Bureau of the Census, Washington, D. C.