

"ONWARDS TOWARDS A 2000 CENSUS DESIGN: RESEARCH RESULTS"

by R.D. Tortora, S.M. Miskura, D.A. Dillman

Discussion by Ivan P. Fellegi, Statistics Canada, 26-A, R.H. Coats Building, Ottawa, Canada

Introduction

In my discussion at the ASA session last year on the 2000 Census, I suggested to the Census Bureau that they urgently sort out the basic options that should be tested in 1995 -- as distinct from more distant (although possibly promising) possibilities on which on-going research, and perhaps testing in the year 2000, should be carried out. A great deal of progress has occurred since, but I still have some concerns. I would like to start by re-stating a few basic background facts.

Background facts

Census data are used in two fundamental ways. First, there is the use of the count or estimated population for the constitutional requirement of reapportionment and for the legal requirements of redistricting and fiscal transfers. Second, there is the enormous substantive data base (mostly long form data) which is the most fundamental source of comparable information on small areas and/or small and slowly changing population groups, based on a single overall methodology implemented at a single point in time. Given its constitutional-legal character, the first kind of use is more fundamental and certainly more public. Nevertheless, I venture to assert that most substantive users' interest relates to the second kind of use.

Keeping in mind these two classes of census use, I support the view that the two most important failings and/or problems of the last several censuses have been the inability to reduce the differential undercount, and the dramatic increase in costs. The first failing is, of course, directly at conflict with the constitutional-legal objective. The issue of dramatic cost increases is directly endangering the census as a broad, substantive, multivariate data base. The two problems intricately interact. Indeed, I believe that it is the succession of attempts to remedy the problem of differential undercount that has led the Bureau to layer after layer of complexity and cost -- unfortunately without the desired effect. But now the combination of lack of success on this front, together with the high and growing costs, has led to a fundamental questioning of whether the count

related objective could be accomplished using some completely different approach, such as an administrative records based census, at much lower cost and with a differential undercount that is at least not worse. The consequent loss of the multivariate data base that is the census as we know it would be a terrible, almost irreplaceable one for the much more numerous second class of users. As it happens, there is a growing consensus, at least among experts (see the National Academy of Sciences' recent report of the National Academy of Sciences Panel on the Year 2000 Census Requirements), that an administrative records based census is not feasible for the year 2000. But this is a reprieve that we must use judiciously to deal with the two fundamental problems mentioned above.

Areas of significant progress

The most visible progress since the paper presented at last year's ASA session is the winnowing down of alternatives to be tested in 1995 -- although it is still not clear how these multiple features will be integrated into a single operation, yet evaluated individually.

It is also positive that the Bureau moved from the consideration of fourteen possible Design Alternatives to the consideration of a single package incorporating what it regards as the best features of each.

There are lots of excellent ideas on which the pre-1995 research has concentrated, particularly in the areas of respondent motivation.

There are two further very significant areas of progress, to both of which I shall come back later, since they are also a source of concern. First, there is an articulation, in Appendix 1, Figure 1 of what are called "Foundation Objectives" -- namely "cost effectiveness", "reduce differential coverage", and "effective content". Second, the Bureau seems to have embraced the idea of trying to overcome the persistent problem of differential undercount through statistical estimation -- a very significant step indeed.

Areas of concern

My overall concern is that, while the paper

identifies, correctly in my view, what it regards as the three "Foundation Objectives", there is no indication that these objectives were used rigorously in the decision process leading up to the particular proposed package to be tested. I will use these same three objectives as headings to elaborate my concerns.

(a) Cost effectiveness.

The closely related issues of cost and complexity are not addressed frontally. Indeed, if many of the proposed features designed to enhance coverage and/or to improve response rates are layered on top of the 1990 machinery of census taking, the costs may well increase; the logistic difficulties certainly will.

This is not to say that some of the test features will not likely reduce costs, at least compared to what they might otherwise have been. For example the experiments leading to improved response rates may well have a beneficial effect on costs. But a frontal attack seems to be missing. Let me be specific. The 1970 Census cost \$700 million in 1990 dollars. If one increases this cost in proportion to the growth of the number of housing units, one arrives at a figure of \$1,030 million. Taking into account the declining response rates since 1970 (a figure estimated for the 1990 Census operations as \$17 million for each percentage point), we arrive at \$1,250 million as the nominal cost for the 1990 Census using the 1970 general methodology. But the 1990 Census cost more than twice this much. What is missing is a systematic examination of the cost-benefit ratio of every feature of the 1990 Census that is different from the methodology of 1970. This is all the more needed since probably many of the high cost and high complexity features of the last two censuses were motivated by the overwhelming desire to reduce the differential undercount. These costs might be less justified in the year 2000 census in which an important share of the burden to resolve the differential undercount will be carried by estimation methods.

Even in terms of what is to be tested or developed, it is not clear that cost-benefit considerations have been dominant. For example maintaining inter-censally a continuously updated

Master Address File may well be justified for reasons that have little to do with the census. But in terms of the strict requirements for a 2000 Census the question at least has to be posed, and answered: how does it contribute to the "Foundation Objectives"? It might marginally reduce differential coverage errors -- but the main burden for that objective will now presumably be carried by estimation, since all other approaches of the last thirty years have in retrospect proven to be only marginally productive. Of course, if it makes a major contribution to the reduction of the differential undercoverage, or to a significantly improved population distribution by small areas, then much higher costs could be justified. But the question at least should be posed and answered empirically.

The so-called administrative records based census is another important question related to cost-effectiveness. I personally share the view expressed by the National Academy of Sciences panel that an administrative records based census is not feasible for the year 2000. Nevertheless, there is a belief in some important quarters that it would provide a cost-effective solution. It might be quite important to deal with the issue by trying to put together for the 1995 test sites a composite file derived from administrative records for comparison with the test census file. This could also be the first step towards incorporating into the year 2000 census a larger scale test of the idea. Finally, this could give a boost for a program of exploitation of administrative records for the development of a more active program of small area inter-censal estimates -- an objective that I personally regard as realistic and desirable.

So, in summary, while the paper presents ideas for the **addition** of beneficial features to the methodology of the next census, it does not provide evidence of a fundamental, zero-based review of **existing** features which might have outlived their usefulness. Until that is done, I am afraid that the risk posed for future censuses by concerns over high costs will not be lifted.

(b) Reducing differential coverage.

Although not stated explicitly, the approach to a "one number" census appears to signal a fundamental change in strategy: to let estimation carry the main burden of reducing the differential undercount. This, I believe, is a major step forward. But here too I have

a concern. A "one number" census might be too heavy a burden to carry for estimation. There has been practically no research on the impact of undercount adjustment on any other census output than the counts and distributions by age, sex and race-ethnicity. Would not a more modest objective be more likely to prove achievable -- such as the one to come up with good estimates of the age-sex-race distributions through adjustment, leaving all other census characteristics estimated through the traditional census process? My concern is both for the integrity of the multivariate census data base, and for the setting of realistic objectives with respect to the problem of differential coverage.

(c) Effective content.

I already mentioned under the previous heading my first concern about the multivariate data base aspects of the 2000 census: the unknown effect on the long form questions of the "one number" census objective. Another concern relates to the possibility of abandoning altogether the follow-up of non-respondents. Follow-up on a sample basis I can understand; no follow-up (or at least no follow-up beyond a certain point in time) is difficult to accept.

Matrix sampling is another feature about whose benefits I am doubtful. Mostly as a result of layers upon layers of methodology added on top of one another, the logistics (and consequently the cost) of censuses has increased substantially. Matrix sampling would add another major complexity. For what benefit? The most important blocks of questions on the long form need to be cross-tabulated with one another: labour market information, industry and occupation, education, income, race-ethnicity. If most of the corresponding questions therefore have to end up on all variants of the different long form questionnaires, then the differences among them will boil down to carrying different blocks of lower priority questions. In effect, the gain would be to accommodate more lower priority questions. Does this justify a significantly larger cost and complexity?

Concluding comments

The paper represents a very significant, indeed in some ways dramatic, evolution in the thinking of the Census Bureau. I believe that several of the points I have raised have more to do with the

presentation than with the thinking. However, as a friend of not only the Census Bureau but censuses in general, I firmly believe that it is incumbent on us to address systematically all the "foundation objectives" identified by the Census Bureau and to render explicit how every design and testing initiative is **directly** related to the achievement of these objectives. We are, collectively, under a magnifying glass -- indeed, we are on probation.