# CONTINUOUS MEASUREMENT ALTERNATIVES TO DECENNIAL ENUMERATION

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

**Introduction**

I am supposed to discuss the excellent paper by Keith Rust. But, in effect, we are both discussing the Census Bureau's current plans for research on continuous measurement (CM for brevity), as outlined in various papers by Charles Alexander. Keith has mostly put the positive side of what is, indeed, an interesting proposal. I will try to assess it more critically.

I will focus on two strategic questions. First, is CM a realistic and cost effective alternative to the long form for the year 2000 Census? And, second, looking at the full breadth of potential requirements on the statistical system **beyond 2000**, should CM be given a central priority?

The year 2000 Census

In planning for the year 2000 Census, two strategic concerns surfaced: the persistent differential undercount of blacks and Hispanics, and cost. Continuous measurement received its recent impetus from congressional belief that dropping the long form from the census would materially contribute to resolving both of theses concerns. Let us examine the evidence.

As far as coverage is concerned, the coverage error caused by missing a **complete households** is primarily due to errors in the address register, not to the type of questionnaire mailed to a household that is known to exist. And coverage errors **within households** are, according to Census Bureau data, just about the same for households whether they received the short or the long form. At any rate, as Keith noted in his paper,even if some small coverage improvement could be obtained by dropping the long form, its effect would be minimal since it would apply to only one sixth of the households.

Concerning costs, the Census Bureau is not proposing cost savings by replacing the long form with CM -- it is only offering, at best, a cost neutral alternative. At any rate, no cost bargain here!

So CM itself makes no contribution to the two key strategic concerns about the 1980 and 1990 Censuses. But both of them can be addressed within the new context created by the Census Bureau's plan to produce a so-called "one number census" in the year 2000. This would involve, during its last stages, statistical techniques to estimate and distribute the number of those left uncounted up to that point. That these techniques directly address the issue of differential undercount is evident -- that is their explicit objective. It is less evident but just as important that the "one number census" represents a new reality for census taking within which costs should be reducible by a very significant margin -- but only if this becomes a new focus for a concerted effort to redesign the census process itself.

Indeed, in the 20 years between 1970 and 1990 the per household cost of the census, in constant dollars, increased by almost 140% -- while the questionnaire length remained unchanged. Some of the cost increase was undoubtedly associated with generally more difficult social conditions. But I believe that a substantial share was due to a relentless pursuit of the key objective of materially reducing the differential undercount. In spite of best efforts, however, the differential undercount persisted, leading the Census Bureau to the "one number census" design in which the final stage of the census uses statistical techniques to correct for the undercount which persisted up to that point. In this new context there is a strategic opportunity to re-engineer the front end of the census. This should include, but go beyond, the elimination of those operations whose contribution to coverage is marginal relative to their cost. In order to provide a focus for such re-engineering, it might be desirable to establish an ambitious cost saving objective, e.g. to design a year 2000 Census whose unit cost per household is "only" let us say 50% higher than that of the 1970 Census. While the effort devoted to CM does not directly compete with this kind of re-engineering, in a world of limited resources the opportunity costs might be important.

So CM does not contribute to dealing with the two strategic challenges facing the year 2000 Census -- indeed it competes with them for resources. Let me turn now to the issue of its presumed cost neutrality

with the census long form. The proposed monthly sample size for the so-called intercensal long form is 250 thousand households. Over ten years this is more or less equivalent to **twice** the sample size of the long form. Leaving details aside, is cost neutrality believable in a macro sense? On the one side is CM, a stand-alone survey spread over ten years, and with an overhead that needs to be carried over that entire period. On the other side is the long form which is entirely piggy-backed on a census operation that at any rate has to collect the short form information from the every households. All my prior expectation is that under these conditions the unit cost of CM simply has to be higher than that of the long form. Yet cost neutrality would require that continuous measurement should have a per household unit cost that is **half** the cost of the long form! Indeed, Census Bureau simulations show that the savings in 1990 from dropping the long form would have been about $3-500 million. The current decade-long cost estimate for CM is about $600 million. Even this undoubtedly low number exceeds the savings from dropping the short form by some an amount in the range of $1-300 million. This shortfall is supposed to be made up by largely uncosted and unspecified savings elsewhere in the system. It seems to me that proponents of CM should either be required to provide reasonable evidence for cost neutrality or, more realistically, they should explicitly justify the additional budget required for CM involved in terms of its benefits compared to the long form.

Even these cost comparisons might be too favourable to CM since they are based on simulated savings in 1990. But the long form operation in 1990 involved an intensive attempt to follow up every non-respondent household. Current plans call for a reduced level of follow-up intensity for the 2000 Census. Consequently the cost of the long form, hence the scope for savings from dropping it, will be significantly reduced.

This is still not all. Existing legislation requires the Census Bureau to provide small area data, specifically from the census, on education, place of birth, citizenship, year of entry of immigrants, language, and most importantly on income. If these legislative requirements are not changed, the Census Bureau will be **required by law** to collect these variables in the census, on a perhaps abbreviated long form. So, unless the relevant legislation is changed, the potential savings from dropping other long form variables would further diminish, probably close to insignificance.

Apart from its comparison with the long form on the basis of cost and as a source of small area data, there are some other, system-wide implications of continuous measurement to consider as well. For example, the Current Population Survey and other ongoing surveys provide some absolutely critical data that, at the national level, would differ from the lower quality estimates generated by CM -- potentially with a monthly frequency as well. What is the proposed resolution of, for example, conflicting data on unemployment? More fundamentally, through what approach, using what mechanism, and testing over what period of time would these issues be addressed? Presumably, this would have to happen before the critical decision on substituting CM for the long form.

This brings me directly to my final point about the 2000 Census: the issue of critical dates. According to current plans the first substantial test of CM is to commence in early 1997. Evidence having to do with costs, small area data performance, and user acceptance must await at least a year of data, followed by at least several months of evaluation. So one might expect the dust to **begin** to settle in late 1998. But early in 1997 the Census Bureau must seek congressional approval for the list of topics to be collected in the 2000 Census, presumably including whatever is or is not proposed for a long form.

So I clearly believe that CM, while interesting, is not viable as an alternative to the 2000 Census long form, that it is certainly not a cost neutral alternative, and that the substantial research interest that is focused on it risks diverting attention from the two strategic issues of public and congressional interest: cost and differential undercoverage. What about its potential role beyond 2000?

Beyond the year 2000

Looking beyond 2000, I want to make three points having to do with strategic priorities.

First, our societies face formidable challenges requiring statistical data to help guide fundamental policy changes and to monitor their impact. Just to illustrate: our health system will undoubtedly be thoroughly restructured, but we do not know nearly enough about the real impact on either life expectancies or on the quality of life of various health

interventions. Indeed, we know far too little about the determinants of good health -- let alone about the possible cost effectiveness of alternative approaches to it. Similar comments apply to our knowledge of education and life long training, and their potential impact, for example on success in the labour market. And we lack the necessary understanding of the workings of that tangled web that we loosely call social safety nets. I believe that we, as statisticians should lead a determined campaign to develop vitally important information on these issues. Funding is clearly a prerequisite, as is a quite unprecedented mobilization of our best efforts. By comparison CM, with its net decade long cost of several hundred million dollars, is interesting and useful: it would provide more frequent intercensal small area data on traditional topics. But, in competition with other unmet statistical needs, it would not be my first priority if I had a few hundred million dollars to expend on the statistical system.

Second, even within the context of meeting intercensal small area data needs, I would give much higher priority to the relatively inexpensive and high pay-off work of further exploiting administrative records. This is particularly so since it is my understanding that there is no groundswell of client demands for the type of more complex data that CM would be able to generate.

Finally, there are some important conceptual and analytic problems to be resolved. Let me mention just a few for purposes of illustration. How does one compare the five-year average of two municipalities, one of which was growing strongly, the other declining strongly during the period? What is the analytic meaning of median income when the income data for some people in the moving average refers to the beginning, some to the end of the averaging period? In the same vein, what is the analytic interpretation of comparisons of poverty rates for small areas computed by using continuous measurement?

Conclusion

My questions are mostly **not** addressed to Keith Rust whose paper I was supposed to discuss. Most of them are not even addressed to Charles Alexander, since they primarily concern not technical issues but rather statistical system priorities involving opportunity costs. Undoubtedly, there **are** questions about the plausibility of available cost estimates and other assumptions involving continuous measurement. No doubt, further resources and testing can resolve these -- though not in time for the year 2000 Census. But, for now, the implausibility and uncertainty are part of the current context within which priority management decisions must be made about where to place our energies so as to foster the most productive evolution of the statistical system. Perhaps it is the most important problem of all that it is not clear: who, if anyone, has the unambiguous authority within the statistical system to deal with these system-wide issues.