DISCUSSION

Leslie Kish, University of Michigan

Looking over all three papers concerned with stratification, let me congratulate the six authors for bringing us again good news from the Bureau of the Census. The three papers exhibit fine, technical, statistical skills. These skills combine statistical theory with practical understanding and put both to work on real problems. They also involve developing or modifying sophisticated computing programs and efficient utilization of the excellent high speed computers that the Bureau can happily provide.

The three papers deal with methods for improved stratification for PSU's. This is an important aspect of sample design, because the gains from stratification are commonly much greater for between PSU components than for within components or for element sampling. Reasons for that contrast have been given before, are known to samplers, but this view is still neglected in most publications on stratification.

Furthermore, leadership by the Bureau is especially welcome in this field, because the Bureau has the best resources in technical manpower, in computing facilities, and in data banks. The Bureau also has the greatest motivation here, because of potential benefits to its own large samples from research in this field. But the benefits of this research are also shared by smaller survey centers, hence the Bureau's work in this field is a particularly good example of optimal (or proximal) allocation of statistical resources and budgets.

Now let us become more specific, and begin with the Alexander-Kobilarcik paper on new and more specific stratifiers for crime surveys. The Bureau found their results to be assaulted by disturbingly large between PSU components on key variables of crime surveys. These large between components survive the reductions brought by stratification of PSU's with the common geographic, urbanization and similar variables, which are more effective for multipurpose surveys like CPS and for the multisubject areas served by the Bureau's master sample of PSU's.

It seems reasonable to expect that specific crime indicators would yield considerable gains as stratifiers. Hence the investigation was entirely justified. The negative results conflict with conventional wisdom. Incidentally, they should also serve as caution to those who would rely (or tell us to rely) recklessly on models that they cannot check.

We need not regret altogether the negative results. They help us avoid or neglect two types of problems that would be emphasized by strong positive results, which would face us if the specific crime stratifiers would have proved much better. These would be problems not only specifically for the Census Bureau but for other survey samplers as well. First, it would make a multipurpose design, especially a multisubject selection of PSU's, seem much less efficient than sets of PSU's specifically designed and operated for single purposes. However a common and continuing utilization of one sample of PSU's yields great economies, and these present justifiable sources of inertia to changes of PSU's for each new subject. Thus the negative news are welcome and useful for continued use of master samples of PSU's not only in the Bureau but also in other centers.

Second, the test was made essentially on the survey variables (crimes), rather than on the kind of variables usually available for stratification. Variables available for stratification usually do not have very high correlations with survey variables. We generally use several of these moderate stratifiers and it is good to hear that such multivariate stratification compares well with stratifiers that are merely hypothetical and unavailable. More of this later.

The harmful effects of the large PSU components on the Bureau's crime survey are mitigated because they affect only about that third of the entire sample which does not come from self-representing areas. Perhaps the effect could be further mitigated with larger samples of PSU's, perhaps based on a modified or changed field operation, even perhaps covered by travelling interviewers. The large PSU effects may be permanent features due to haphazard, unpredictable causes; such as the widespread depredations in single counties of one or a few individuals or gangs - until they become deactivated. This may be tested by investigating the variability over years of high crime rates in counties.

I welcome also the motivation and thrust of the papers on clustering algorithms by Judkins and Singh and by Kostanich and her partners. I welcome new methods for multivariate stratification, new techniques utilizing the capacities of modern computers, and quantitative, empirical assessments of their results. I particularly welcome more flexible modes of stratification, and add that neither Cochran or I were devotees of "rectangular stratification". I have never practiced it, and I have often talked and published against it; for example, "Objectivity and Regularity Unnecessary" is the title of section 3.6E in my book. Though I have not used clustering methods for actual survey samples, I have followed the attempts, since the 1945 paper by Hagood and Bernert on component indexes, to utilize large numbers of stratifying variables by combining them. Clustering methods appear increasingly attractive as algorithms become more sophisticated, as computing programs become more accessible, and as more data for stratification become available. As L.J. Savage would say wryly: here is a powerful new tool looking for uses.

Nevertheless, before we accept these methods as contributions to applications for use in actual samples, they must make progress in four directions, at least. First of all, their gains must be measured not by how well they cluster the stratifying variables that go into their composition, but by how much they reduce variances for survey statistics. There are no sufficient reasons to believe that superior performances in the former yield strong inferences for superiority in the latter. Even variables with the same name have different values in different years. These remarks are similar to those I made earlier on the other paper. They are reinforced by the lower gains shown in Tables 1-4 than in Table 5 of the Kostanich paper.

Second, the comparisons for gains should be shown not only against unstratified samples, but chiefly against good stratification by traditional

(but flexible) techniques. Traditional methods have virtues in meaningful, understandable categories, and in pliability to changes. We found (Kish and Anderson, JASA 1978) that principal components did not perform better than multivariate stratification, or even as well.

Third, and very important, the comparisons should be highly multivariable, to fit the multipurpose and multisubject utilization of the stratified samples of PSU's.

Fourth, the comparisons should include not only the global sample, but also the many domains that the multipurpose samples must satisfy. Reducing variances for important domains may even be more important than for the large samples of relatively precise global estimates (see the paper by Lorah et.al.).

To be frank, I must say that these first results as they stand here, without the four further steps I indicate, can do more harm than good if anybody accepts them as guides to design. And if it is bad for practice, it is not good theory either, I say.

One minor suggestion for the Kostanich paper: it would be easier to compare results on single charts showing the gains (or better still the remaining variances) of all five techniques against each other. And another for the Judkins-Singh paper: it is possible to restratify and still retain most PSU's (see Kish and Scott, JASA 1971).