United Nations Expert Group Meeting to
Review Critical Issues Relevant to the Planning of
the 2010 Round of Population and Housing Censuses
15-17 September 2004
New York

A proposal under study as part of the planning of the 2010 Mexican Population and Housing Census*

By

Alfredo Bustos
Instituto Nacional de Estadística Geografía e Informática
(INEGI)

I.   INRODUCTION

1.   The most important objective of every census is to provide decision makers in government, as well as in other areas, with an accurate description of the situation presented by some characteristics of a certain population at a given time. When the population is formed by people, the relevant characteristics are, among others, demographic, educational or occupational; on the other hand, when the population being studied is formed by production units belonging to one or more economic sectors, the topics include production, employment or sales.

2.   Recently, a number of alternative proposals have been put forward in order to tackle the growing conflict which arises between the above mentioned objective and the costs involved in fulfilling them. Good examples can be found on the Internet sites of the American Community Survey,[1] of the UK Office for National Statistics (ONS)[2] or of France INSEE.[3] Furthermore, national statistical systems (NSS's) are being asked to produce more frequent and geographically more detailed information on a greater number of topics, which only adds to the stress imposed on them and on their costs of operation, when everything else, methodologies included, remains the same.

3.   In Mexico, INEGI has been considering the suggested alternatives as well as their costs and the required information infrastructure (e.g., population and/or household registers). Legal requirements, like the formation of electoral districts on the basis of the population information provided by the most recent census, have also entered into this analysis. As a result of these considerations, the conclusion that, in the long run, the system of national socio-demographic statistics has to be revamped relying more heavily on the use of administrative records, has been reached. Many such records, however, which have to be developed, installed and running for efficient implementation of some of the above mentioned alternatives, are unlikely to be ready in Mexico by 2010. Therefore, the suggestion has been made that, in the meantime, the relationship between Information Quantity and Quality (IQQ) and cost (C) be given a more careful consideration.

4.   The basic idea is that the mere elimination of one or more parts of the collection and analysis strategy to achieve a reasonable reduction of costs does not necessarily result in an improvement of the IQQ-to-C relationship. Complementing a simple-form census with the collection of further information via surveys adds flexibility and frequency to the results but may negatively affect comparability between small geographical areas. In any case, it seems that most of the methodologies suggested rely on an initial (and, in some cases, periodic) universal headcount of individuals and/or dwellings plus additional information on them, required to define reasonable strata for sampling purposes.

---

[1] http://www.census.gov/acs/www/
[2] http://www.statistics.gov.uk/events/RSS_ONS_Conf/presentations.asp
[3] http://www.insee.fr/fr/recensement/page_accueil_rp.htm

II.   MEXICO'S CENSUS EXPERIENCE IN THE 2000 CENSUS ROUND

5.   During the planning stages of the 2000 Mexican Census, information users were approached in order to gain some insight as to their requirements. A careful revision of Mexico's national and international statistical commitments was also carried out. The results were, as expected, the identification of a large number of highly relevant topics for decision makers in the public and private sectors as well as in academia.

6.   However, should those requirements be expressed as batteries of questions, and these in turn, reflected in paper forms, the result would be both prohibitively expensive and inefficient to collect; inefficiencies would arise in view of a heavy burden imposed on the respondent. The obvious conclusion is that it is impossible and useless to ask everything to everyone in a heterogeneous population of nearly 100 million.

7.   Users were then asked to review their expectations in the light of the above considerations. Even though considerable reductions were achieved, the situation was not very different from the one described above but with an added element: user's needs were not fully met. In other words, census planners were faced with a no-win situation: high costs – poor quality – diminished analysis possibilities.
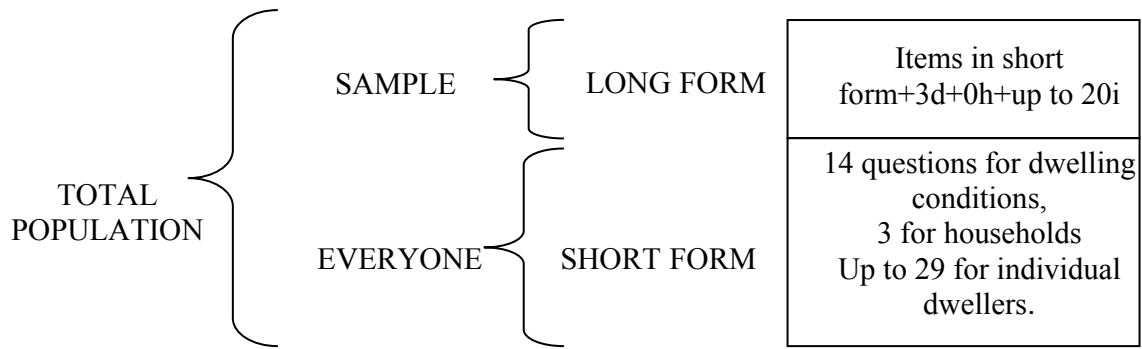
8.   Under these conditions a number of options were put forward, among which the following two were considered:

a. To reduce the number of topics and items included in the census form. This would on the one hand, guarantee lower costs, and consequently better manageability, as well as better quality of census results. On the other, such gains would be obtained at the expense of diminished analysis capabilities for some topics and the total elimination of others. One advantage of this approach, however, would be that the information that does get collected would refer to geographical units of all sizes completely covering the Mexican territory.
b. To resort to a short- long-form strategy. While costs would be marginally larger and quality would suffer a little, a greater number of topics could be looked at in greater depth. Since the long form would be given to a sample, its results would not apply to small geographical areas. In other words, a compromise between geographical representation and subject matter coverage would be called for. However, once again, user's expectations would be far from being fully met.

A. Collection forms

9.   Alternative b above was finally implemented in 2000.[4]

---

[4] The two census forms used, may be found at:
http://www.inegi.gob.mx/est/contenidos/espanol/proyectos/censos/cpv2000/cuestionario/imagenes/basico.pdf
http://www.inegi.gob.mx/est/contenidos/espanol/proyectos/censos/cpv2000/cuestionario/imagenes/amplio.pdf

			Items in short form+3d+0h+up to 20i
TOTAL POPULATION	SAMPLE	LONG FORM	
	EVERYONE	SHORT FORM	14 questions for dwelling conditions, 3 for households Up to 29 for individual dwellers.

B. Diagnosis

a. Analysis capabilities

10. It has already been noted that under the current methodologies, census exercises cannot cover in number and in depth all topics suggested by users. This is mainly due, but not exclusively, to budgetary constraints. Even within a two format strategy the additional questions introduced in the long form cannot be expected to satisfy fully users' stated needs.

b. Geographic detail

11. Though limited, the topics covered by universal enumeration may refer to small geographic areas and provide a fair description of the conditions within its borders. The progressive accumulation on the basis of spatial criteria provides fairly accurate ideas as to the behavior of a population in wider areas.

12. The situation, however, is less favorable for those topics covered through sampling. Indeed, some descriptive measures available under different conditions are either impossible to produce or have a high degree of uncertainty which renders them useless.

c. Dissemination

13. Traditionally, more widely disseminated census results have been limited to tables and graphs showing some descriptive statistics (e. g., totals, means and ratios) on the basis of questions answered by the whole population. In turn, long-form results receive the same treatment given to most survey studies. But despite the greater amount of information they offer, they too are limited to descriptive tables and graphs.

14. In other words, short- and long-form results are treated independently from each other. In fact, the advantages of performing parallel exercises are not fully used. It is as if neither has anything to gain from the other.

d. Conclusion

15. Under the census methodologies currently in use in Mexico and other countries, based mostly on a descriptive approach to statistics, it is unlikely that the relationship IQQ-to-C
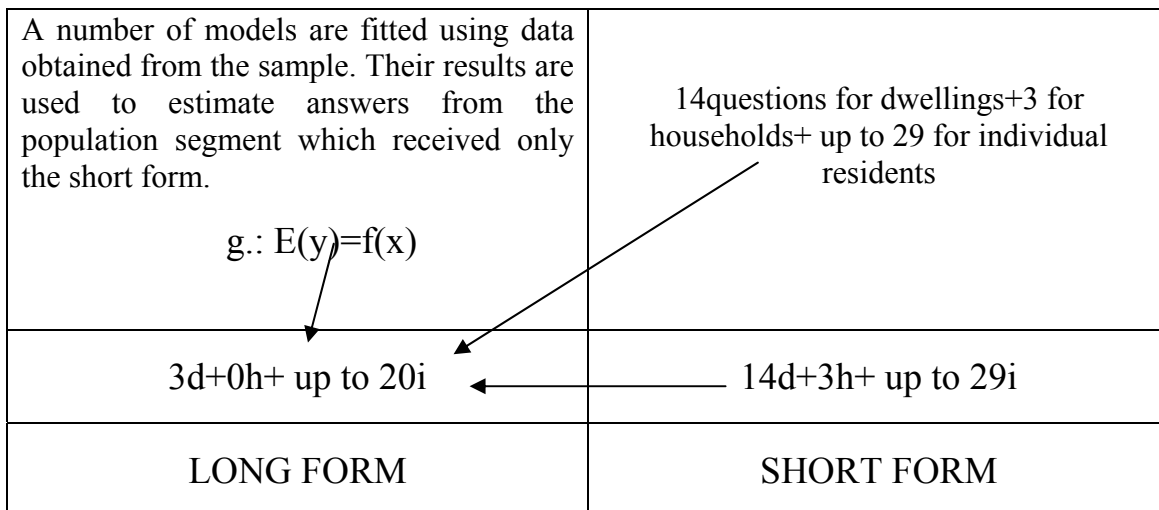
can be improved. It seems necessary to explore the introduction of some alternatives not available until recently in order to achieve improvements in the above relationship. In particular, mathematical statistics has kept on producing new results which can and should be applied in the present context.

16. Thus, it can be stated that the potential of the two-form strategy for the 2000 Mexican census has not been fully exploited. Statistical modeling and its application in estimation can help to extend the results already attained by complementing individual records with estimates. Should the degree of uncertainty surrounding such estimates prove too large, the modeling and estimation of household-, block- or county-aggregates could be attempted. Some questions to be addressed are:

   a. Can the relationships between items in the short and long forms be explicitly stated?
   b. Can they be adequately estimated from the long form data only using already linked records?
   c. Can the latter set of answers be estimated from the former to a given degree of precision?
   d. If this proves to be the case, is it reasonable to use the resulting estimates to impute the answers to items present only in the long form to those in the short form?
   e. Is it more sensible to follow this approach for local aggregates?
   f. Finally, and most importantly, does the above imply improvements on the products we seek and, therefore, in the knowledge users will derive from census exercises (IQQ-to-C)?

17. This line of reasoning, which is summarized in Figure 1 below, is already being put to the test using the year 2000 census database.

Figure. 1

A number of models are fitted using data obtained from the sample. Their results are used to estimate answers from the population segment which received only the short form. g.: $E(y)=f(x)$	14questions for dwellings+3 for households+ up to 29 for individual residents
3d+0h+ up to 20i	14d+3h+ up to 29i
LONG FORM	SHORT FORM

18. Clearly, the above exercise does not completely address the noted limitations; in particular, in terms of the number of topics and depth with which they are studied.

However, even if only partially successful, some gains to be obtained from it can already be noted:

  a. Detail. The suggested estimation for small geographical areas will enhance the amount of information made available to the user at little extra cost. Thus, a positive impact on IQQ-to-C.
  b. Analysis. Studies based on those estimates will allow policy-makers to zero in on target populations.
  c. Infrastructure. Development of methodologies and software which will allow similar analyses to be performed even under different circumstances.
  d. Expertise. Development of personnel who will carry out similar analyses even under different circumstances.
  e. Proposals directed at improving methodologies.

III.   PROPOSAL FOR 2010

19. Pending the results of the above exercise, and as a consequence of the previous discussion, a third proposal, additional to the two considered for the 2000 census, has been put forward in order to overcome the remaining shortcomings, for the year 2010 census in Mexico. This proposal is the following:

20. To develop a short form to be used with the vast majority of the population and, in addition, to design as many census forms as may be necessary in order to increase the number of topics treated as well as the depth with which they are covered. Each form in the second group would be used on a separate sample. All the samples would be designed so that no individual household gets more than one form.

21. The information thus collected would be used to fit all the necessary models so that estimates could be obtained at different geographical levels of aggregation depending on their predictive capabilities. The number of forms could be reduced whenever it is feasible to include two or more topics in one form.

22. It is expected that budgetary requirements will be substantial anyway, since it is unlikely that any of the cost-cutting options under discussion be implemented by 2010, and that option 3 will result in small increases. It is also expected that improvements in quality and quantity of the results will more than offset them. The availability of information at the local level will improve; analysis capabilities will be widely enhanced in view of the increase in number of topics dealt with.

23. It must be stressed that the proposal under discussion should not be seen merely as a "transition alternative" for two main reasons: (1) other alternatives require an initial headcount; but for budgetary constraints, it is up to us to make them as informative as we possibly can so that later actions be solidly anchored, and (2) all of the proposals for alternative methodologies will require, at one stage or another, the use of estimation techniques from survey data; therefore, whatever the decision made by the relevant institution, they will find the availability of the infrastructure and the expertise developed to be to their advantage.

A. Strategy

24. It is necessary to develop a careful strategy so that users of census results in the year 2010 have at their disposal geographically detailed information covering a large number of topics in greater depth and that is thus relevant, timely, comparable, precise, accessible and sufficient. To this end, it is necessary to be prepared on a number of fronts.

a. Issues and topics

25. As new collection options become available, it is useful to revise and, wherever possible, enrich the set of topics and questions to be considered in a census. In particular, it is important to reach consensus as to their current relevance in the national as well as in the international context. It is therefore necessary to consider the following actions, among others:

- Meetings with experts.
- Agreement on those subjects whose treatment by a census is inadequate.

b. Study of theoretical and statistical relationships among variables

26. Since each census form is to be developed from scratch, the opportunity arises for its optimization. From the study of the relationships among variables, one could think of reorganizing the basic set of questions (i. e., the short form) so as to include in it those with better predictive power. In order to establish such relationships it is useful to begin by reviewing the literature on the subject matter. Of course, the results of such a review have to be compared with the experience in each country and thus validated. The information gathered by statistical agencies in previous censuses or via surveys becomes a valuable input. The analysis is subject to theoretical constraints given that the information was not gathered from a designed experiment with this purpose in mind.

c. Design of census forms

27. In theory, at least, it is possible to assume that different sets of variables have different predictive power for variables in the complementary set. Whether an optimal partition can be determined is yet to be seen. The analysis of previous statistical experience may help, however, to achieve a reasonable approximation or to reject altogether such idea. If the proposal has any merit, the idea of statistical criteria being used as an aid in the design of questionnaires is tempting.

d. Sampling design

28. Requirements such as the one that no household receive more than one form or those which have to do with the estimation process imply the need for a careful design of each sub-sample. For instance, if only block aggregates are desired, it is not clear whether only one form should be used in all sampling units within the same block or a combination of them would be called for in order to gain more information.
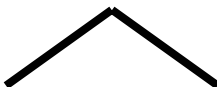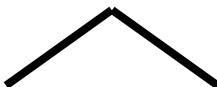
e.   Final consensus for Mexico

29. It is expected that when the previous sequence is iterated a couple of times, enough experience will be available so as to reach adequate consensus among users and producers. Such consensus will materialize in contents and structures for the design of census forms to be used during the collection stage of the 2010 population and housing census in Mexico.

B. Requirements

30. Little more is required other than the strengthening of institutional capabilities, by means of the analysis of statistical bibliography, a group of statisticians with solid background in statistical modeling, analysis of available commercial software, a fairly robust computer infrastructure and time to perform these activities.

Figure. 2

COMMON QUESTIONS	+ QUESTIONS IN TOPICAL FORMS				
Questions and topics with greater predictive capabilities or unpredictable, applied to the whole population	**Sample 1: Income**	⌃	⌃	. . .	⌃
	⌃	**Sample 2: Fertility**	⌃	. . .	⌃
	. . .	. . .	**Sample 3: Migration**	. . .	
	. . .	. . .	. . .	. . .	. . .
	Sample n: Income				**Sample n: Violence**