

Henry Woltman, U.S. Bureau of the Census

I. Introduction

The purpose of this paper is to provide a general overview of the current status of the 1980 Census content error evaluation program. For this purpose, "content error" is defined as an error of observation or objective measurement, of recording, of imputation or other processing which results in associating a wrong value of the characteristic with a specified population element - a housing unit or person in our case. Errors due to undercoverage of the population or housing units are excluded.

As in past census evaluation programs, the general objective of the content error evaluation program is to provide estimates of the nonsampling variance and bias components of estimates derived from the 1980 Census. These components include the simple response variance, the correlated response variance and response bias. Briefly, the simple response variance is the basic trial-to-trial variability in response for a given population element averaged over all population elements; the correlated response variance component is the contribution to the total response variance from the correlation among response deviations (i.e. the difference between the individual reported value and the average value over trials under the same general conditions) introduced by the interviewer, the editor, the crew-leader, the coder etc. Finally the response bias for a population element is the difference between the average reported value over trials for the element and the true value, and the net response bias for an estimated total is the average of the individual response biases.

The following sections of this paper will focus on a discussion of the plans developed to date to measure simple response variance and response bias associated with various population and housing data items included on the 1980 Census questionnaire. First I will outline the general nature of various content error evaluation techniques considered for use in this part of the evaluation program and subsequently provide more detailed discussion of the allocation of specific 1980 census data items to one or more of these techniques.

A study to measure the correlated response variance component associated with various census operations identified as potential major sources of nonsampling error in 1980 Census data is also being considered. A further stated objective of this study is to provide estimates of the total mean square error of various census statistics. At present, plans for this study are undergoing review and a presentation of the details of this study, its specific objectives, scope and methodology, will have to await the results of this review.

II. Proposed Content Error Evaluation Methodologies

The following techniques have been considered for use in evaluating simple response variance and response bias in the 1980 Census. Most of these have been used in previous Census content error evaluation programs and as such represent the traditional approaches to evaluating these nonsampling error components.

A. Reinterview

This evaluation technique consists of selecting a sample of households enumerated in the Census and conducting a second interview at the household. The results are matched to the Census at the address level to evaluate the reported characteristics of the housing data items and at the person level to evaluate the reported personal data items.

The reinterview can be designed to either (1) repeat the census question, or a format similar to the census question using to the extent possible, the Census procedures or (2) administer, through personal visit interviews and self response, a series of detailed probing questions designed to help the respondent better understand the intent of the census question or to enable the respondent to better organize and recall the information desired.

In the former case, a measure of simple response variance is obtained which is likely to be an under-estimate due to trial-to-trial conditioning of the respondent. In the latter case a measure of response bias or more accurately a measure of the differential response bias is obtained if we consider the reinterview as providing a more accurate response, on the average. In this case, one can view the reinterview as measuring a characteristic using the "preferred" measurement technique as compared to the "working" technique necessitated for use on the census questionnaires (both short and long form) due to space limitations, overall respondent burden considerations and other constraints imposed by the census data collection procedures.

B. Matching Studies to Current Sample Surveys

Sample surveys conducted on a continuing basis which coincide with the census enumeration time frame and which measure data items included on the census questionnaire provide another means by which to evaluate content error. Two surveys are considered as prime candidates for matching to the census - the March 1980 Current Population Survey (CPS) to evaluate population data items and the Annual

Housing Survey (AHS) conducted during October-December 1980 to evaluate housing unit data items. The Current Population is a monthly survey conducted to collect data on labor force participation while the Annual Housing Survey has been conducted yearly since 1973 and collects data on changes in the housing inventory as well as detailed characteristics of the housing inventory. For example, utilization characteristics (e.g., rooms, bedrooms), structural and plumbing characteristics, equipment and fuels characteristics and financial characteristics are collected. The data for these surveys are generally collected by personal visit or in some cases by telephone, after an initial personal visit, using a household respondent. These matching studies provide the opportunity, although admittedly imperfect, to evaluate simple response variance, or perhaps what might more properly be termed response consistency, considering that we are comparing the responses obtained using somewhat different data collection procedures. The design of such a matching study may also include a third contact with a household to reconcile response differences between the census and the survey in which case a measure of differential response bias may be obtained. Reconciliation of response differences for some items is considered desirable for the AHS-CENSUS match study due primarily to the time lag between census enumeration and the AHS interview, some six to eight months.

C. Record Check Studies

These studies are designed to provide a more accurate measure of response bias by matching housing units or persons to administrative records which contain presumably "error free" data. Record checks may be designed to evaluate the full range of responses to the census question by selecting a sample from the census and matching to the record source or to evaluate only false negative responses by selecting a sample from the record source and matching to the Census.

D. Other Special Studies

Special studies may be required for evaluation of specific data items due to special requirements on the sample design and/or evaluation methodology that are not met by one of the above studies. For example, special sample designs may be required to evaluate content error for "rare" items with sufficient precision or to evaluate data items where, for example, the preferred survey technique involves written tests, or some such other

specialized measurement technique.

II. Proposed Content Error Evaluation Methodology by Census Item

The content error methodology proposed, if any, for each 1980 Census data item was developed in consultation with the subject matter specialists and is shown in figure 1. In general, a data item was proposed for evaluation based on (1) whether it is a new data item, (2) whether substantive changes in the question format have been made since 1970 and (3) the current applicability of previous census content error evaluations or 1980 pretest content error evaluation data. In figure 1 data items which have undergone substantive revision (i.e. major changes in target population and/or question format) since the 1970 Census are asterisked.

A brief description of the revision is given in attachment A. New data items are devoted by capital N in figure 1 and the specific question sequences to be used are shown in attachment B.

The revised population data item questions are those for Spanish/Hispanic origin, state of birth, disability, relationship, race, and income. For the housing data items, revisions have been made to the questions on plumbing facilities, tenure, utility costs and vacancy status. New questions cover language speaking ability, ancestry, journey to work (time, mode and carpooling arrangements), trucks and vans available for use by household members, and shelter costs for owner occupied units (i.e. costs for taxes, insurance, and mortgage).

Almost all of the data items to be evaluated have been allocated to the content reinterview or to the current survey - census match studies. Several general observations concerning the overall nature of these studies are as follows.

First, the reinterview study will be designed to provide estimates of differential response bias between the measurement procedure used in the reinterview and that used in the Census for population and housing data items which are new or have been substantially revised since the 1970 Census. The revised questions to be evaluated are those on race, relationship, Spanish/Hispanic origin, state of birth, disability,¹ plumbing facilities and vacancy status; the new questions are those for ancestry, language-speaking ability, journey to work, and mortgage status for owner-occupied units (i.e., whether the unit is currently mortgaged or not mortgaged). Present plans are to use a sample of about 14,000 designated household of which about 10,000 are expected to be occupied and reinterviewed. It is also anticipated that some set of primary sampling units

Footnote

¹/ As of this date, development of a large-scale sample survey to measure the characteristics of the disabled and nondisabled populations is being considered. If this survey is funded, it will serve as the evaluation mode for this data item.

selected for the Current Population Survey will be used as the primary units for the content reinterview. This will allow us to make use of current survey interviewers to administer the reinterview questionnaire after receiving appropriate training. It is planned to conduct the reinterview during July-September 1980.

A final methodology issue currently being addressed is that of developing the measurement technique to be used in the reinterview for several data items. The evaluation of the question on Spanish/Hispanic origin is a case in point of the difficulty this presents for some items due to the lack of a precise conceptual definition. The question on Spanish/Hispanic origin is essentially one of self-perception as the question asks "is this person of Spanish/Hispanic origin or descent?" Thus, the respondent's answer to the question is, in many instances, a measure of whether or not a person perceives himself/herself as of Spanish Hispanic origin or descent since the question provides no objective criteria by which an individual's origin or descent is to be determined. In the 1970 Census and in the pretests conducted to develop the 1980 Census questionnaire, ancestry has been used to evaluate the responses to this question. The use of a detailed series of questions on family ancestry to evaluate this question is not entirely appropriate, however, unless one assumes this to be the unstated criteria upon which respondents are to base their answer. In addition, Spanish surname lists have also been used for this purpose although, again, this is not error free since a Spanish surname does not always imply Spanish origin. With the aid of the subject matter specialists, we are currently developing the question sequences to be used in the reinterview.

Second, the Current Population Survey—Census Match Study will provide measures on the consistency of reporting primarily for labor force participation items, occupation and industry classification and educational attainment. Other items common to the CPS and Census such as age, race and relationship will also be evaluated. We expect that this study will yield some 10,000 households interviewed in the March 1980 CPS and matched to a census long-form questionnaire.

Lastly, the Annual Housing Survey—Census Match Study will provide measures of response consistency or possibly differential response bias (in the event response differences are reconciled) for virtually all the housing data items. We also expect that this study will yield some 10,000 households interviewed in the AHS and matched to a census long-form questionnaire.

In addition to these major studies, record checks are proposed to evaluate the value of home and average monthly utility costs (gas and electricity) items and a special study proposed to evaluate the language-speaking ability data item. A brief discussion of the proposed evaluation strategies for these data items follows.

A. Value of Home

For the value of home question we propose to repeat a study conducted as part of the 1970 Census evaluation program. The basic methodology of this study is to select a subsample of property transfers identified in the 1982 Census of Governments Property Values Survey (PVS) and compare the reported sales price of the property with the value reported for that same address in the census. The PVS includes a validation procedure to insure that the sales price reflects a true cash market value. The chief problems with this study are the short time span covered by the property transfers (probably six months) and the time lag of some two years between the census and the PVS. The impact of the latter problem on estimates of response bias may be reduced somewhat by adjusting the 1982 sales price to account for inflation in the housing market. This study is important due to the dramatic changes which have occurred in the housing market since 1970 and as such we have no idea as to the impact such rapid change may have on the respondents ability to accurately estimate "home value" at any particular point in time.

B. Utility Costs (Gas and Electricity)

For the 1980 Census, average monthly costs for gas and electricity will be collected for both renter-occupied units and owner-occupied units. A 1970 Census evaluation and some recent evaluations in the Special Censuses of Travis and Oakland indicate that estimates of the average monthly cost of gas and electricity are subject to relatively large response biases (net over-reporting) and that the size of the bias varies considerably from area to area. The impact of these errors is dampened somewhat when combined with other items to form gross rent and shelter cost statistics.^{2/} However, for certain types of housing units, e.g., owner-occupied without a mortgage, the effect of errors in utility cost estimates is more pronounced. Also, as utility costs increase, these errors may have more impact on gross rent and shelter cost estimates.

Previous evaluation studies have provided information on the magnitude of errors but they do not provide much information on methods for reducing errors. Since many utility companies have computerized billing systems it has been suggested

Footnote

^{2/} Gross rent for renter occupied units equals the value of contract rent plus any extra average monthly costs for utilities. The shelter cost for owner occupied units equals the value of any mortgage payment plus costs for real estate taxes, fire and hazard insurance, and utilities.

that a way of reducing error in reported utility costs would be to have the utility companies furnish their customers with the average amount of their bills for the 12 months prior to the census and include this average cost with the most recent utility bill provided prior to Census Day. This plan would make prior arrangements with utility companies to retain utility costs for their customers for the 12 months prior to the census. Most companies keep records for at least 12 months as a matter of policy. At the time of the census (April 1980) companies would include in the current month's bill a statement of average costs for the preceding 12 months.

In order to evaluate whether or not such a procedure would reduce response error within a geographic area served by a particular utility company, some method of distributing the average utility cost information will have to be devised so that comparisons of census reports could be made for those household receiving average cost information with a control group of household not receiving the information.

C. Language Speaking Ability

As shown on attachment B, the language question asks whether or not each person speaks a language other than English, and if so what language and how well the person speaks English - very well, well, not well or not at all. It is clear that the last part of the question is very subjective in nature and as such it is questionable whether it will provide meaningful data. As a possible evaluation technique we are considering the use of a language proficiency test developed for the Department of Health, Education and Welfare to determine the ability of a person to speak and understand English.

IV. Concluding Remarks

To date, the broad outlines of the content error evaluation program have been developed but the specifics of the sample design and sample size for many of the individual studies proposed, are still unknown. With regard to the design of this phase of the total evaluation program, it is important to recognize that the proposed content error studies, along with the studies proposed for coverage evaluation and experimental census procedures, are all competing for the same space, staff, and money. Over the next few months, we will be making decisions about which of these studies will be undertaken and, for those to be undertaken, addressing the methodological and sample design issues in greater detail.

V. Acknowledgements

The development of this phase of the 1980 Census evaluation program represents a joint effort of many persons within the Bureau of Census. In particular the author would like to recognize the contributions of Larry Love (Field Division), Dave Koons and Betty Kent (Housing Division), Paula Schnieder and Mark Littman (Population Division) and Larry Carstensen (Statistical Methods Division).

Figure 1
Proposed Content Error Evaluation Studies By Data Item

Data Item	Reinterview	CPS Census Match	AHS Census Match	Record Check	Special Study	No Evaluation Proposed
<u>Population Items</u>						
*Relationship	X	X				
Sex						X
*Race	X	X				
Age						X
Marital Status						X
*Spanish/Hispanic Origin	X					
School Enrollment	X					
Education Attainment	X	X				
*State of Birth	X					
Naturalization/Immigration						X
(N) Ancestry	X					
(N) Language	X				X	
Residence 5 years ago						X
Activity 5 years ago						X
Veteran Status						X
*Disability	X					
Number of Children Born						X
Multiple Marriages						X
Place of Work					X	
(N) Journey to Work (time, type of vehicle, and pooling arrangement)	X					
Employment Status						X
Industry & Occupation						X
Work Experience in 1979						X
*Income						X

*-data item question revised since 1970 Census, see attachment A

N-new data item for 1980 Census, see attachment B

Figure 1
Proposed Content Error Evaluation Studies By Data Item

Data Item	Reinterview	CPS Census Match	AHS Census Match	Record Check	Special Study	No Evaluation Proposed
<u>Housing Items</u>						
Entry to Living Quarters						X
*Plumbing facilities	X		X			
Number of Rooms			X			
*Tenure (owner/renter)			X			
Value of Home				X		
Contract Rent			X			
Number of Living Quarters at Address			X			
Floors/Elevators						X
Year Built			X			
Source of Water						X
Sewer Connections						X
Heating Equipment			X			
Fuels (heating/water heating/cooking)			X			
*Utility Costs				X		
Kitchen Facilities			X			
Number of Bedrooms			X			
Number of Bathrooms			X			
Telephone						X
Air Conditioning			X			
Autos (available for use by household member)			X			
(N) Trucks & Vans (available for use by household member)			X			
(N) Real Estate Taxes						X
(N) Insurance (fire and hazard)			X			
(N) Mortgage Type & Payment	X		X			
*Vacancy Status	X					

*-data item question revised since 1970 Census, see attachment A

N-new data item for 1980 Census, see attachment B

Substantive Data Item Revisions Since the
1970 Census

Population Items

1. Relationship
The 1970 Census derived household relationship using the "head of household" as the reference person. The 1980 question uses the household member in whose name the home is owned or rented as the reference person. If no such person exists, any household member may use as the reference person.
2. Race
The specific reference to "color or race" has been replaced with the phrase "Is this person _____" and the response is expected to indicate the racial group the person most closely identifies with. The number of categories has been expanded from eight to fourteen.
3. Spanish/Hispanic Origin or Descent
Categories revised. Specific question revised to require a Yes/No response as well as the specific origin (e.g. Mexican, Chicano, Puerto Rican).
4. State of birth
Asks specifically for state or foreign country in which mother was living when person born. The 1970 Census question asked "where was this person born?"
5. Disability
Scope of question expanded from work disability to also include use of public transportation and to be asked of persons 65 years old and over.
6. Income
Section on sources of income from other than Wage and Salary and self employment income expanded. A total income figure for 1979 is also requested.

Housing Items

1. Plumbing Facilities
Consolidates the threepart 1970 Census question on hot and cold piped water, flush toilet and bathtub or shower into one question.
2. Tenure
Category to report unit as a cooperative or condominium which is owned or being bought eliminated. New question on whether or not apartment or house is part of a condominium added. No count of cooperatively owned units will be obtained.
3. Utility Costs
(electricity, gas, water, other fuels)
Expanded to include both owners and renters. The 1970 Census question applied only to renters.
4. Vacancy Status
In addition to vacancy status and months vacant the question includes a new part to identify "boarded up" units.

NEW QUESTIONS IN THE 1980 CENSUS

I. Population Data Items

A. LANGUAGE SPEAKING ABILITY

- Replaces 1970 Census question on mother tongue

13a. Does this person speak a language other than English at home?
 Yes No, only speaks English — Skip to 14

b. What is this language?

 (For example — Chinese, Italian, Spanish, etc.)

c. How well does this person speak English?
 Very well Not well
 Well Not at all

B. ANCESTRY

- Replaces 1970 Census questions on parents' country of birth.

14. What is this person's ancestry? *If uncertain about how to report ancestry, see instruction guide.*

(For example: Afro-Amer., English, French, German, Honduran, Hungarian, Irish, Italian, Jamaican, Korean, Lebanese, Mexican, Nigerian, Polish, Ukrainian, Venezuelan, etc.)

II. Housing Data Items

A. TRUCKS AND VANS

H29. How many vans or trucks of one-ton capacity or less are kept at home for use by members of your household?

None	2 vans or trucks
1 van or truck	3 or more vans or trucks

B. HOUSING COSTS FOR ONE-FAMILY UNITS

Please answer H30–H32 if you live in a one-family house which you own or are buying, unless this is —

- A mobile home or trailer
 - A house on 10 or more acres
 - A condominium unit
 - A house with a commercial establishment or medical office on the property
- If any of these, or if you rent your unit or this is a multi-family structure, skip H30 to H32 and turn to page 6.*

H30. What were the real estate taxes on this property last year?

\$ _____ .00 OR None

H31. What is the annual premium for fire and hazard insurance on this property?

\$ _____ .00 OR None

H32a. Do you have a mortgage, deed of trust, contract to purchase, or similar debt on this property?

- Yes, mortgage, deed of trust, or similar debt
- Yes, contract to purchase
- No — Skip to page 6

b. Do you have a second or junior mortgage on this property?

Yes No

C. JOURNEY TO WORK

- Parts a, c, and d are new. Part b. modified to obtain mode "usually" used last week.

24a. Last week, how long did it usually take this person to get from home to work (one way)?

Minutes

b. How did this person usually get to work last week? *If this person used more than one method, give the one usually used for most of the distance.*

- Car Taxicab
- Truck Motorcycle
- Van Bicycle
- Bus or streetcar Walked only
- Railroad Worked at home
- Subway or elevated Other — Specify _____

If car, truck, or van in 24b, go to 24c. Otherwise, skip to 28.

c. When going to work last week, did this person usually —

- Drive alone — Skip to 28
- Drive others only
- Share driving
- Ride as passenger only

d. How many people, including this person, usually rode to work in the car, truck, or van last week?

- 2 4 6
- 3 5 7 or more

After answering 24d, skip to 28.

c. How much is your total regular monthly payment to the lender?

Also include payments on a contract to purchase and to lenders holding second or junior mortgages on this property.

\$ _____ .00 OR No regular payment required — Skip to page 6

d. Does your regular monthly payment (amount entered in H32c) include payments for real estate taxes on this property?

- Yes, taxes included in payment
- No, taxes paid separately or taxes not required

e. Does your regular monthly payment (amount entered in H32c) include payments for fire and hazard insurance on this property?

- Yes, insurance included in payment
- No, insurance paid separately or no insurance

Please turn to page 6