

Janet A. Tippett, Bureau of the Census

INTRODUCTION

The characteristics of the housing inventory are collected from two major sources by the Census Bureau: the decennial census (and its attendant pretests) and the American Housing Survey (AHS) [1]. The first census of housing was in 1940, and some comparisons with other data sources were made on an occasional basis, especially with the Housing Vacancy Survey (HVS) data collected as part of the Current Population Survey during the 60's and early 70's. However, it was not until the advent of the American Housing Survey (AHS) in the fall of 1973 that substantive housing characteristics survey data were available for comparison with decennial census results. The AHS, as a longitudinal survey, began to provide intercensal data to reflect the dynamics of changes in the housing inventory, both nationally and for selected metropolitan areas. Initially, the survey provided updates to 1970 "benchmark" data.

Since the 1980 census results would present the opportunity for a comparison of decennial and sample survey housing data, an AHS-Census match, with partial reconciliation, was planned as part of the 1980 evaluation program, but had to be curtailed for budgetary reasons. In mid 1983, a small-scale AHS-Census Match Pretest was designed, as a feasibility study of a large evaluation study. Results of that study (see [2]) indicated it is feasible to match, reinterview and reconcile, but that timing of the reconciliation is very important. Even at best, it is a check of the reliability of respondent reporting and lacks an independent estimate of the accuracy ("truth") of the characteristics being reported. It is quite possible, for example, for the same household to report consistently to both census and survey that their house was built in the 1950's, whereas in fact the actual date was in 1947. Consequently, we began to look for independent sources (administrative records) to provide a 3-way comparison. The 1985 test census of Tampa provided just such an opportunity.

DESCRIPTION OF THE TWO DATA SOURCES

The census and the AHS have methodological differences of note. While the census produces both 100% and sample data, collected primarily by self-enumeration, the AHS is a sample survey conducted via personal and telephone interview, generally by experienced interviewers. The content of the census shows great continuity over time, defining as it does basic or "benchmark" data. AHS consists of longitudinally maintained "core" questions, many of which are identical or similar to census questions, but also allow for a broader subject area and for a more experimental questionnaire format.

The advantages of the census are a large sample size, a previously tested questionnaire format, and a representative and comprehensive set of respondents, typical of all types of housing units and their occupants. The AHS, with

an interviewer present in a half-hour (average) session, presents an opportunity to answer respondent questions, and should result in fewer "don't knows" or blank items from respondents. However, several studies of surveys and interviewer performance have indicated the possibility of enumerator bias, leading the respondent, or making entries without asking all questions exactly as intended.

RATIONALE FOR MATCHING

In spite of methodological differences, or perhaps because of them, comparing answers to identical or very similar questions from the same household provides insights into the advantages and limitations of each method.

In addition, data users frequently use the sources jointly or in a complementary fashion to analyze the housing inventory, especially on the local level. The more we know of the quality of the data, including non-sampling as well as sampling error, the more effective will be our responses to data user needs. Where an independent record source exists, an additional evaluation of data quality becomes possible. The Tampa census test provided an excellent opportunity, since it was the site of a 1990 test census, fell into the AHS metropolitan sample for the first time in 1985, and had available records for certain housing characteristics of note.

METHODS OF EVALUATING COMPARATIVE DATA

The commonly accepted method of evaluating data from either census or survey is through content reinterview, conducted, preferably independently, as soon as practicable after the original interview. This method effectively measures the variance, and sometimes the bias, of the data. When reconciliation is part of the reinterview, some real insights into the quality (and limitations) of the data emerge. While this provides a consistency check and thus focuses on the reliability of the data, it does not, however, totally indicate whether the data effectively measure the characteristics of interest. The latter is best measured by use of a good independent data source which collects or provides the same information for the same units (or households).

RECORD CHECK

Definition. An administrative record includes any record maintained by a person or by a public or private organization which identifies a set of persons, addresses and/or companies which are participants, licensees or recipients of a prescribed activity. Under this definition, membership lists would be included as well as public records such as license applications, tax assessments, lists of program recipients, and census survey records. Each record contains a set of variables, which may include: (1) name, or address or some other identifier; (2) economic

or demographic information about each universe member (for example, age, race or sex of household members, or property value or characteristics of their housing).

Advantages/What to Look for in Administrative Records. The major advantage of an independent record source is to contribute to "truth" or to provide a way of evaluating whether the data we produce accurately reflect "real-world" concerns of credibility as well as meeting statistical standards.

In considering whether a record check is appropriate, one needs to consider what I call the "4 A's" of Administrative Records:

Availability. Do records exist for the population (universe) or locality of interest? Are they automated? What is the format? Would it be feasible to do a match?

Access. By this I mean access to the office or person who controls the records to determine the possibility of accessing the records themselves. It also entails who would access (personnel), how (computer terminal or clerical search) and where (at the Bureau or at the file location). Would the records be provided gratis, or are there costs involved in accessing?

Accuracy. How good are the records? How long have they been kept? For what purpose? How current are they? What are their limitations?

Adaptability. Having determined the answers to (1) through (3), how suitable are the records for the project's purpose? If the format varies somewhat, how difficult would it be to adapt? For example, if a data file is indexed by account number, could it be readily rearranged by address list?

THE TAMPA STUDY

Although reinterview (and reconciliation) was employed on the Tampa project, the focus of this paper is on the items selected for a record check. The census portion of the data was provided by photocopies of the questionnaires secured between March and July of 1985. A mail-out, mail-back procedure was employed, and about 58% of the households responded (self-enumerated). The remaining 42% were secured through enumerator follow-up, by telephone or in the field by personal visit. During the same time, the city of Tampa was being enumerated for the first time by an AHS interviewer. These interviews took place between April and November 1985. A list of the sample units was obtained and matched against those households who received long form questionnaires (1 in 6 households). For matched households, a comparison of 33 housing characteristics was made. Two items, "year built" and "value" of homeowner properties, were selected for a record check.

THE "4A's OF TAMPA

Availability. In early 1985 telephone inquiries to the city of Tampa indicated that records of interest were maintained for both "year built" for both owners and renters, and "value" for homeowner properties in the Hillsborough County Tax Assessor's Office in downtown Tampa.

Access. An initial personal visit by the project manager took place the last week of

April, to meet with the staff of the assessor's office. At that time it was discovered that "value" data were available on microfiche, whereas "year built" data were kept on assessor's property cards in a separate storage room. Both records were accessed by folio number, which could be obtained by a lengthy map and plat book search [3], beginning with the property address. A chance remark by one of the Tampa staff revealed that if the owner's name as well as address were known, the folio number could be obtained directly from an alphabetical list.

Accuracy. Records for both items are updated annually, which was important for "value" but not for "year built". Records were collected by assessors and had been maintained for years for most city property, including the oldest units. An adjustment factor had to be applied to the assessed value, as Tampa has a \$25,000 homeowner tax exclusion, if the owner lives on the property and applies for it. For most properties, both assessed and market value were shown.

Adaptability. Given the information obtained in the initial site visit, a second visit of the project manager and 3 staff members took place the first week of November. Using two alphabetized lists, one by name for owners, and one by street address for renters, as many cases as could be searched during 5 working days were completed, and the data transcribed from the records directly to the address printouts (containing addresses from the AHS-Census Match). A third list of multi-unit rental properties, used in another Tampa project, was also searched, to the degree time permitted. [4]

RESULTS

Year Built. Analysis of data from the 1980 census illustrates the effect of misreporting by structure age cohorts. A standard assumption is that each decade cohort of housing would decrease in size as it is reported in subsequent censuses due to losses from demolition, disaster, and various conversions to non-housing unit use or to fewer units through merger. Components of Housing Inventory Change surveys support this assumption. They have shown that additions to the housing inventory by various types of conversions or through structural rehabilitation have been far less than losses and that new construction is the predominant source of addition to the housing universe. However, the expected decrease in housing stock does not occur consistently for national data and this variation is even pronounced for Tampa.

Results of the search and record check are shown in Table 1. Table 1 presents an aggregate summary of "year built" data for the 3 data sources.

For 88 units (of 194) all 3 data sources were in agreement (46%).

The census agreed with the assessor's file for 110 of 194 units (incl. 88 above) (57%).

AHS agreed with the assessor's file for 123 of 194 units (incl. 88 above) (64%).

Neither source was in agreement with the assessor's file for 49 of 194 units (26%); for 23 of these (47%) AHS and census agreed with each other (consistency, but not "truth"). For 8 of the census cases and 15 of AHS, the item was a blank (or don't know).

Table 1.--Comparison of "Year Built" Data:
Three Sources

Intervals	Census	AHS	Assessor's Files
1980 or later	4	3	7
1970-79	29	37	34
1960-69	47	49	40
1950-59	48	49	61
1940-49	22	19	29
1939 or earlier	<u>21</u>	<u>22</u>	<u>23</u>
Subtotal	171	179	194
Blank/Don't Know	23	15	0
Total	194	194	194

Case-by-case comparison tables.--Excludes blanks/don't knows .

Census owners.--Of 153 owner-occupied units with an observation in both data sets, 128 (84%) were in agreement.

AHS owners.--Of 119 cases, 96 (81%) were in agreement.

Census renters.--Of 63 cases, 46 (73%) were in agreement.

AHS renters.--Of 59 cases, 52 (89%) were in agreement. Results are a little misleading since there were more blanks/don't knows (10) than for census.

Multi-unit structure respondent.--Of 105 units, 97 (93%) were in agreement.

Conclusions. For "year built" data, all three sources were in agreement for 46% of the total cases (incl. blanks) (Table 1). The census agreed with the assessors' file for 57% of the cases; AHS agreed with the assessor's for 64% of the cases.

"Year built" was better reported by census owners than renters.

Census: Owners-assessors 84%
Renters-assessors 73%
AHS : Owners-assessors 81%
Renters-assessors 89%

The structure respondent (see [3]) proved the best source of "year built" data--93% agreement with the assessor's file.

Value. The "specified owner-occupied" value universe in both Census and AHS is restricted to single-family units on a property of less than 10 acres, which does not contain a medical or business office. Condominiums and mobile homes are excluded. Table 2 presents a comparison of the 3 sources.

Table 2 presents an aggregate summary of "value" data for the 3 data sources.

For 21 units (of 125), all data sources were in agreement (17%).

The census agreed with the assessor's file for 48 of 125 units (incl. 21 above) (39%).

AHS agreed with the assessor's file for 35 of 125 units (28%). There were large numbers of item blanks (25 for AHS and 3 for census).

Case-by-case comparison tables.--Excludes blanks/don't knows.

Census spec. owner-occupied.--Of 122 units with an observation in both data sets, 48 (40%) were in agreement.

AHS spec. owner-occupied.--Of 100 units with an observation in both data sets, 35 (35%) were in agreement.

Table 2.--Comparison of "Value" Data:
Three Sources

Intervals	Census	AHS	Assessor's Files
< \$20,000	8	4	4
20,000-29,999	8	8	5
30,000-39,999	21	8	13
40,000-49,999	33	17	24
50,000-59,999	19	21	34
60,000-79,999	16	29	28
80,000-99,999	6	4	8
100,000-149,999	4	8	3
150,000-199,999	3	2	3
200,000 or more	4	3	4
Subtotal	122	100	125
Not reported	3	25	0
Total	<u>125</u>	<u>125</u>	<u>125</u>

Note that Tampa has a homeowners tax exclusion law, whereby owner-occupants who apply annually are not taxed on the first \$25,000 of property value. It is possible that a number of respondents gave us this reduced figure rather than the full market value.

Conclusions. For "value" data, all three sources agreed for only 21 of 125 cases (17%) (Table 2). The census agreed with the assessor's file for 39% of the cases (including those where all 3 were in agreement). AHS agreed with the assessor's file for 28% of the cases.

There was a 21% item non-response rate for AHS (unedited data).

Viewed independently, the census respondent report for value (from an interval choice display) agreed with the assessor's file in 40% of the matched cases. AHS respondents (giving a point estimate which was converted to the census intervals for comparative purposes) agree with the assessors file in 35% of the cases. However, for 39% of the census cases and 49% of the AHS cases, responses were within 1 interval of the assessor's records.

For a high percentage of cases (26% of units with an assessor's response) the assessor's file agreed with neither data source. (This figure includes the 3 census blanks and 25 AHS blanks.) For many of these census and AHS were in agreement, (23 of 49) indicating those households were reporting value consistently, but inaccurately. A major reason for this may be homeowners' confusion over market value (requested in both census and survey) vs assessed value. In Tampa, owner-occupants may apply annually for a tax exclusion of the first \$25,000 of their property value. It is quite conceivable that this reduced amount was inaccurately reported as market value.

LIMITATIONS OF THE DATA

A number of study limitations should be noted, particularly as they relate to the data sources. In an actual decennial census, much more attention is given to following up and completing long form questionnaires for missing information. In the Tampa test census, a higher percentage of missing data occurred than is customary. The AHS source was an unedited data printout supplied shortly after initial keying of the question-

naires. Actual published tabulations will be edited and missing data allocated. There was a high item non-response for value in AHS (21% of completed interviews).

The last intended panel of AHS interviews was curtailed for budgetary reasons, reducing the intended sample from 500 to 428 cases. The results described in this paper should not be interpreted as indicative of "year built" and "value" reporting nationwide, or for other areas.

The time allowed for searching the records in the assessor's office was limited to a single work week (4 persons for 5 days). Because of the tedious process of locating folio numbers for renters, the list of owners was more completely researched than was the list of renters.

As noted above, the "value" results for Tampa may be affected by that city's homeowners tax exclusion. Furthermore, even though a standard checklist is used by assessors in establishing value, there is a degree of subjectivity in that item not found in "year built".

RECOMMENDATIONS

The study results confirmed some existing hypotheses and indicate areas for further research efforts:

Owners tend to be better informed about the year their housing was built than do renters. There is not so much a reluctance to answer this item, as that respondents simply do not have the information to answer accurately and are giving their "best guess". For multi-unit rental units, an informed structure respondent (see [5]) was a much better source. Consequently, it is recommended that further structure respondent

testing be done in other locations.

The Tampa results indicate problems with the accuracy of respondents' reports of value. A record check is therefore a very important adjunct to the customary reinterview evaluation, which primarily measures consistency. Because of the particular homeowner tax exclusion in Tampa, it is recommended that further administrative record checks be undertaken in other cities (areas).

ACKNOWLEDGMENTS

The author wishes to acknowledge the peer reviews furnished by Bill Mockovak and Pat Kelley of the Census Bureau, as well as the comments of the discussant, Gary Shapiro, also of the Census Bureau.

NOTES AND REFERENCES

- [1] Formerly entitled The Annual Housing Survey.
- [2] Tippett, J. and Takei, R., "AHS-Census Match Pretest", 1980 Census Preliminary Results Memorandum No. 100, July 22, 1985.
- [3] A plat book is a large bound volume containing diagrammatic records of building lot and block number, and a folio number for referencing.
- [4] In a separate Tampa Project, data on multi-unit structures (rental units) were obtained from knowledgeable sources such as property manager, realtor, resident manager, etc. "Year built" was one of the data items obtained from these sources.
- [5] See [4].