

COVERAGE OF HOUSING IN THE 1990 DECENNIAL CENSUS

Danny R. Childers*

Bureau of the Census, Washington, DC 20233

KEYWORDS: undercount, dual-system estimates, coverage evaluation, housing coverage

1.0 Background

The Housing Unit Coverage Study (HUCS) was designed to evaluate the coverage of housing units in the 1990 Decennial Census. The coverage of housing units in the census is important to users of the housing data. Housing unit coverage also impacts person coverage. The persons living in housing units had less chance of being captured in the census if the housing unit was not included in the census address files.

The major objective of this study was to provide evaluation data from the 1990 census for planning the next census. The Census Bureau has proposed to continuously update the list of addresses from the 1990 Decennial Census to support future programs, including the 2000 Decennial Census. This evaluation of coverage of housing units in the file of 1990 census addresses will help to identify portions of the file needing extra improvements before the next census.

The HUCS sample was half of the housing units sampled in the Post Enumeration Survey (PES) (Hogan 1992). The HUCS sample consisted of two parts, the P-sample and the E-sample. The P-sample was an independent listing of housing units in the sample blocks. The E-sample was the housing units enumerated in the census in the same sample of blocks. The P-sample and the E-sample for HUCS were overlapping samples of approximately 80,000 housing units. The P-sample estimated the gross percentage of housing unit omissions within the census search area. The E-sample estimated the gross percentage of housing units erroneously enumerated in the census in the search area. For matching we defined the search area to be a predetermined area for searching in the census for matches to P-sample housing units and for assigning correct or erroneous enumeration to the census E-sample housing units.

2.0 Dual-System Estimates

The Housing Unit Coverage Study was designed to produce an estimate of the net coverage of housing units within each post-stratum in the form of a dual-system estimate. The dual-system estimates were computed for the 180 post-strata. The post-strata were defined by:

Census Region, Place Type, Size of Structure, and Occupancy/Tenure Status. The four census regions were Northeast, South, Midwest, and West. The three types of place were large urban (i.e., 250,000 population or larger), other urban, and rural. The five classifications of size of structure were single unit structure, small multi-unit structure (i.e., 2 to 9 housing units), medium multi-unit structure (i.e., 10 to 49 housing units), large multi-unit structure (i.e., 50 or more housing units), and other structures (i.e., mobile homes, tents, vans, boats, etc.). The three categories of occupancy/tenure status were occupied by an owner, occupied by a renter, and vacant.

For details of the dual-system estimator see Wolter, 1986 and Hogan, 1993. Mathematically, the dual-system estimator is written

$$DSE = \frac{N_p}{M} * (N_c - (EE * \frac{N_c}{N_e}))$$

where

DSE = the dual-system estimate of the number of housing units.

N_c = the census count of housing units.

N_p = the weighted P-sample estimate of the number of housing units.

N_e = the weighted E-sample estimate of the number of housing units.

M = the weighted P-sample estimate of the number of matched housing units.

EE = the weighted E-sample estimate of the number of erroneous enumerations.

3.0 Percent Undercount

The dual-system estimates of the housing units in the 180 post-strata were added to calculate an estimate of the total number of housing units. The percent net undercount is estimated by

$$\text{Percent Net Undercount} = 100 * (1 - \frac{N_c}{DSE})$$

The estimates of standard error in this paper were design-based stratified jackknife estimates computed using VPLX, a general-purpose variance estimation software package developed by Robert E. Fay, Senior Mathematical Statistician at the Census Bureau. For more information about VPLX see Fay, 1990. The estimates of standard error in all tables are in parentheses. All hypothesis tests were at a significance level of 5 percent. A multiple comparison methodology was not employed for the hypothesis testing.

* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

3.0.1 Gross Omissions and Erroneous Enumerations

A gross omission does not necessarily mean the housing unit was missed in the census. A P-sample housing unit was classified as an omission when it could not be located in the census within the search area. An E-sample housing unit was classified as an erroneous enumeration when the housing unit should not have been enumerated within the search area.

The objective of this study was to estimate a net undercount using estimates of the housing unit omissions within the search area and the erroneously enumerated housing units within the search area. However, it is instructive to study the gross housing unit omissions and gross erroneous enumerations separately to investigate trends in the census.

3.0.2 Reasons for Erroneous Enumeration

The six reasons for erroneously enumerated housing units were: within block duplicate, surrounding block duplicate, geocoding error, nonexistent as a housing unit, insufficient information for matching and follow-up, and the portion of unresolved cases imputed to be erroneously enumerated.

A within block duplicate was a housing unit enumerated twice in the E-sample within the sample block. A surrounding block duplicate was a housing unit enumerated in the E-sample within the sample block and also enumerated within the search area outside the E-sample. The housing units classified as geocoding errors were housing units enumerated within the sample block that actually existed outside the search area.

The classification of nonexistent as a housing unit contained several types of nonexistent housing units. These erroneous enumerations were census housing unit enumerations that should not have been enumerated as housing units within the search area. For more details of nonexistent as a housing unit see the HUCS Results Memorandum Number 4, 1993.

The classification of insufficient information was defined as an address in the census files without enough information to locate the housing unit for an interview. These housing units had minimal or blank addresses and the maps were missing. The minimum amount of information required for HUCS follow-up was a housing unit spotted on the map for minimal or blank addresses.

A housing unit was coded unresolved when the field follow-up interview for HUCS was inconclusive. The probability of correct enumeration in the census for the housing units coded unresolved was imputed in order to calculate dual-system estimates. A logistic regression model was fit to resolved E-sample housing units to predict the probability of correct enumeration for unresolved housing units. For more details of the imputation strategy as it applies to the Post Enumeration Survey see Belin et al., 1993.

3.1 Occupancy Status

Only the undercount in occupied housing units is of interest to many researchers, since the undercount in vacant housing units does not affect the undercount of persons. The estimated percent undercount for occupied and vacant housing units is presented in Table 1.

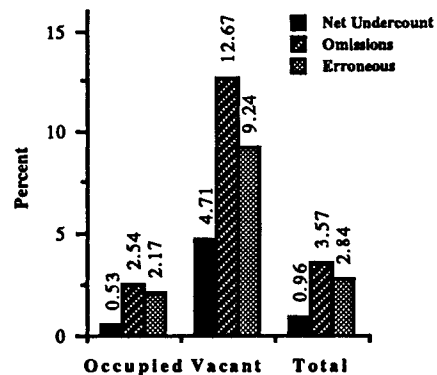
Table 1: Percent Net Undercount for Occupancy Status

<u>Occupancy Status</u>	<u>Percent Undercount</u>	
Occupied	0.53	(0.21)
Vacant	4.71	(1.26)
Total	0.96	(0.24)

The percent net undercount was reduced by almost half when the vacant housing units were ignored. The percent net undercount for the vacant housing units was significantly larger than the percent net undercount for occupied housing units.

The percentage net undercount is compared to the percentage of gross omissions and to the percentage of gross erroneous enumerations for housing units in Figure 1. The percentages of omissions and erroneous enumerations for the vacant housing units were significantly greater than for the occupied housing units.

Figure 1: Occupancy Status



The reasons for erroneous enumeration are compared for occupied and vacant housing units in Table 2.

Table 2: Percentage of Erroneous Enumerations for Occupancy Status

<u>Reasons for Erroneous Enumerations</u>	<u>Occupied</u>	<u>Vacant</u>
Within Block Duplicate	21.8 (2.2)	12.2 (4.4)
Surrounding Duplicate	18.9 (2.4)	4.8 (1.3)
Geocoding Error	22.3 (4.0)	2.8 (0.8)
Nonexistent	24.4 (2.7)	66.0 (5.6)
Insufficient Information	10.6 (2.1)	9.4 (2.7)
Unresolved	2.0 (0.3)	4.8 (1.0)

The occupied and vacant housing units were erroneously enumerated for different reasons. Adding the percent of erroneous enumerations from within and surrounding block duplication, an estimated 40.7 percent of the erroneous enumerations in occupied housing units were duplicated with a standard error of 3.2 percent. An estimated 17.0 percent of the erroneous enumerations in vacant housing units were duplicates with a standard error of 4.6 percent. The percentage of erroneous enumerations that were duplicated for occupied housing units was significantly greater than vacant housing units.

The percentage of erroneously enumerated housing units that were geocoding errors was significantly larger for occupied housing units. The effect of geocoding error on the vacant housing units was almost nonexistent.

The amount of vacant housing units not existing within the search area as housing units was probably influenced by the difficulty in determining "vacant and boarded-up" versus "not fit for habitation". A housing unit may have the windows boarded-up and look unfit for a person to live in the housing unit, but by census definition the housing unit was fit for habitation if it was not exposed to the elements.

3.2 Census Region

Estimates of the percent undercount for the four census regions for total, occupied, and vacant housing units are presented in Table 3.

Table 3: Percent Undercount for Census Region

Region	Total	Occupied	Vacant
Northeast	0.53 (0.52)	0.30 (0.40)	2.74 (3.69)
South	0.80 (0.43)	0.65 (0.38)	1.92 (1.77)
Midwest	1.13 (0.43)	0.47 (0.30)	7.39 (2.96)
West	1.48 (0.58)	0.61 (0.55)	9.22 (1.91)
Total	0.96 (0.24)	0.53 (0.21)	4.71 (1.26)

A comparison of percent net undercount for the four census regions for occupied, vacant, and total housing units indicated no significant difference among regions. Comparing the occupied and vacant housing units for the regions indicated no significant difference in occupancy status for the northeast and south regions. However, the percent undercount for the occupied housing units were significantly different from the vacant housing units in the midwest and in the west regions.

Estimates of the percentage net undercount, omission, and erroneous enumeration for the four census regions are compared for occupied housing units in Figure 2 and for vacant housing units in Figure 3.

The larger net undercounts for vacant housing units in the midwest and west regions appeared to be a reduction in the erroneous enumerations and not due to larger omission rates.

Figure 2: Occupied Housing Units by Census Region

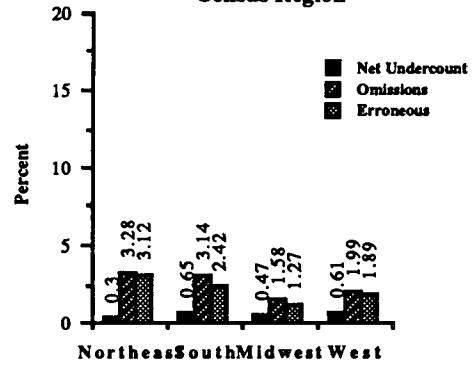
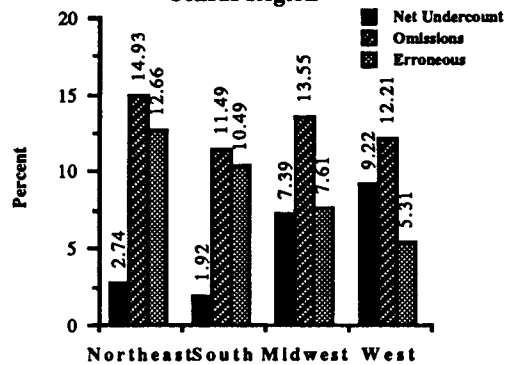


Figure 3: Vacant Housing Units by Census Region



3.3 Type of Place

Estimates of the percent undercount for the three categories of type of place for total, occupied, and vacant housing units are presented in Table 4. The estimated percent undercount for all housing units in the large urban place type was -0.09 percent, suggesting a slight overcount. However, the estimated percent undercount for housing units in large urban were as was not significantly different from zero. The estimated percent undercount for the rural place type was significantly larger than both the large urban and other urban place types for total housing units.

Table 4: Percent Net Undercount for Type of Place

Type of Place	Total	Occupied	Vacant
Large Urban	-0.09 (0.31)	-0.12 (0.29)	0.18 (1.53)
Other Urban	0.91 (0.42)	0.81 (0.39)	1.87 (1.70)
Rural	3.16 (0.53)	1.67 (0.40)	10.21 (2.13)
Total	0.96 (0.24)	0.53 (0.21)	4.71 (1.26)

The address lists for the large urban areas were well defined house number and street name addresses. The Bureau used these address lists obtained from vendors to conduct a mail census. The address lists for the other urban areas were a mixture of vendor lists and lists

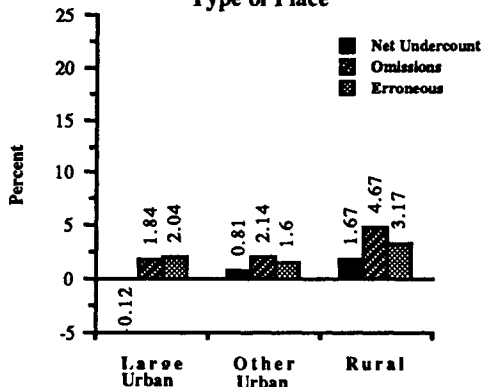
compiled by Bureau personnel. The address lists for rural areas were compiled by Bureau personnel. Many of the addresses in the rural areas were not house number and street name addresses and at times may have consisted only of a location description. These types of rural addresses were more difficult to enumerate. This difference in coverage for the vendor addresses lists and address lists compiled by Bureau personnel may be a function of the type of addresses rather than the method of compiling the addresses.

For occupied housing units, the percent undercount for the rural place type was significantly greater than the large urban place type, but not greater than the other urban place type. For vacant housing units, the percent undercount for the rural place type was significantly greater than both the large urban and other urban place types.

A comparison of percent net undercount in occupied and vacant housing units for the three place types indicated no significant difference in the occupied and vacant housing units in the large urban and the other urban place types. However, the percent undercount for the vacant housing units was significantly greater than the occupied housing units in the rural place type.

Estimates of the percentage net undercount, omission, and erroneous enumeration for the three place types are compared for occupied in Figure 4 and vacant housing units in Figure 5.

Figure 4: Occupied Housing Units by Type of Place

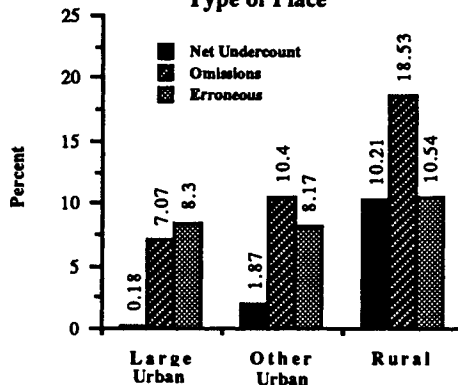


The percent omission for occupied housing units in rural areas was significantly greater than in the large and other urban areas. The percent omission for occupied housing units in other urban areas was significantly greater than large urban areas. The percent erroneous enumeration for occupied housing units in rural areas was significantly greater than in the large and other urban areas.

The percent omission for vacant housing units in rural areas was significantly greater than in the large and other urban areas. The percent omission for vacant housing units in other urban areas was not significantly different from the large urban areas. On the other hand, the percent erroneous enumeration for vacant housing

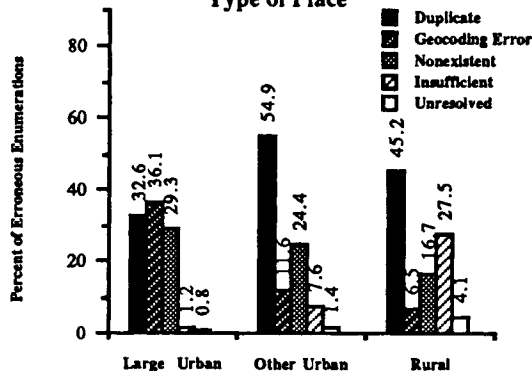
units was not significantly different for the three place types.

Figure 5: Vacant Housing Units by Type of Place



The classifications of erroneous enumeration were examined to better understand the different types of erroneous enumerations for the three place types. The percentage of erroneous enumerations for occupied and vacant are graphed in Figures 6 and 7.

Figure 6: Occupied Housing Units by Type of Place

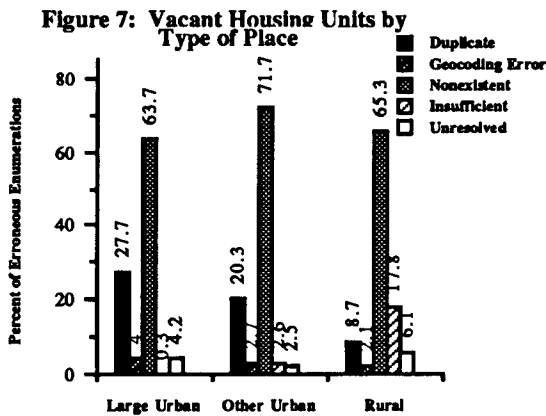


In occupied housing units duplication in other urban and rural was significantly greater than in large urban. The percentage of geocoding error in occupied housing units in large urban areas was significantly greater than the other urban and rural areas.

For occupied housing units in large urban areas duplication, geocoding error and nonexistent contribute to the erroneous enumerations. Housing units with insufficient information and unresolved housing units were rare for occupied housing units in large urban areas.

For occupied housing units in other urban and rural areas duplication was the major reason for erroneous enumeration. Nonexistent housing units were also a factor in the other urban and rural areas. Insufficient information in rural areas was also a contributor to the erroneous enumerations. These rural addresses that can only be located by a map and were vague and at times

incomplete.



The reason for erroneous enumeration for vacant housing units was predominantly nonexistent enumerations for the three place types. Duplication in large urban and other urban place types was also a contributor to the erroneous enumerations. Geocoding error was almost nonexistent for the vacant units in the three place types. The census enumerations with insufficient information did not contribute to the erroneous enumeration in large urban and other urban place types. The census enumerations with insufficient information were a factor for the rural place type.

3.4 Size of Structure

Estimates of the percent undercount for the five categories of size of structure for total, occupied, and vacant housing units are presented in Table 5. The estimated percent undercount for occupied housing units in single unit structures was not significantly different from zero.

The estimated percent undercount for the small multi-unit structures was significantly different from zero. The conversions from large single units to small multi-unit structures sometimes did not appear on mailing lists and many of these housing units did not receive mail separately from other housing units in the structure. The people may even have been counted, if all residents in the structure were enumerated on the census questionnaire delivered to the structure, but many of these persons were also missed.

The estimate of the percent undercount in housing units in medium multi-unit structures was -2.19 percent, indicating an estimated overcount of 2.19 percent. The estimated net undercount was not significantly different from zero for housing units in large multi-unit structures.

The estimated percent undercount for occupied housing units in single unit structures was significantly different from the percent undercount for occupied housing units in small and medium multi-unit structures.

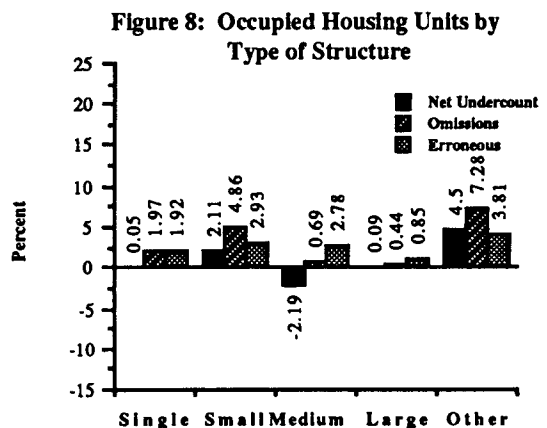
The estimates of percent undercount for the single unit structures were significantly different from the other types of structures (i.e., mobile homes, tents, vans, and boats) for occupied housing units. The mobile homes present problems for census enumeration, because of the ease in moving the housing units. The single and multi-unit structures move rarely and were easier to capture on the Bureau's address files.

A comparison of percent undercount in occupied and vacant housing units for the five categories of size of structure indicated a significant difference in occupancy status in single unit structures. In contrast, the percent undercount for the occupied and vacant housing units were not significantly different for the three sizes of multi-unit structures and the other types of structures.

Table 5: Percent Undercount for Size of Structure

Size of Structure	Total	Occupied	Vacant
Single	0.76 (0.23)	0.05 (0.18)	8.08 (1.54)
Small Multi	2.25 (0.65)	2.11 (0.59)	3.35 (2.43)
Medium Multi	-2.41 (1.22)	-2.19 (1.12)	-3.90 (4.28)
Large Multi	-0.94 (1.23)	0.09 (0.52)	-8.19 (8.37)
Other	4.46 (1.28)	4.50 (1.26)	4.32 (3.81)
Total	0.96 (0.24)	0.53 (0.21)	4.71 (1.26)

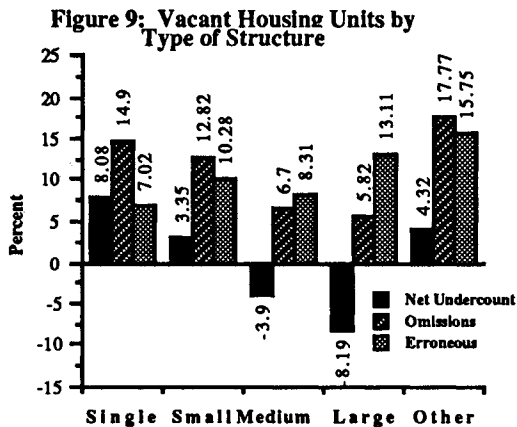
Estimates of the percentage net undercount, not matched, and erroneous enumeration for the five types of structures are compared for occupied in Figure 8 and vacant housing units in Figure 9. The percentage gross omission for occupied housing units in small multi-unit structures was significantly greater than for housing units in single unit structures. Also, the percentage gross erroneous enumeration for occupied housing units in small multi-unit structures was significantly greater than for housing units in single unit structures.



In vacant housing units the percentage gross omission for housing units in single and small multi-unit structures were not significantly different. The

percentage erroneous enumeration for housing units in small multi-unit structures was significantly greater than for housing units in single unit structures.

The percentage gross omission for vacant housing units in small multi-unit structures was significantly greater than for vacant housing units in medium and large multi-unit structures. The percentage gross erroneous enumeration for vacant housing units in small, medium, and large multi-unit structures were not significantly different. In other words, the estimated overcount in medium and large vacant multi-unit structures was not due to an increased rate of erroneous enumeration, but due to reduced omissions of vacant housing units.



3.5 Occupancy and Tenure Status

The estimated percent undercounts for the three categories of occupancy/tenure status are presented in Table 6.

Table 6: Percent Undercount for Occupancy/Tenure Status

Occupancy/ Tenure Status	Percent Undercount	Standard Error
Occupied/Owner	0.37	(0.21)
Occupied/Renter	0.80	(0.39)
Vacant	4.71	(1.26)
Total	0.96	(0.24)

The percent undercount for housing units occupied by persons who own the housing unit was not significantly different from persons who rent the housing unit. The owner and renter status in the tenure variable was an important variable for coverage of persons (Hogan, 1993), but was not significant for coverage of housing units. The higher missed rate for persons who rent the housing unit than for persons who own was probably due to the mobility of the persons within the housing unit. The coverage of the housing unit was not affected by the owner or renter status of the persons living within the housing unit.

4.0 Conclusions

The estimated percent net undercount for all housing units was 1.0 percent. When the vacant housing units were ignored, the net undercount was reduced by half for the occupied housing units. The vacant housing units were missed at a higher rate than occupied housing units.

The net housing unit coverage in large urban areas was extremely good. The undercount in rural areas was larger than in the urban areas. The rural type addresses probably contributed to the housing unit undercount in rural areas. The vacant housing units in rural areas had a larger net undercount than the urban areas.

The net housing unit coverage for occupied single units was extremely good. The difference in net undercount for occupied and vacant housing units was significant only in the single unit structures.

The housing units in multi-unit structures with 2 to 9 housing units and in the other structures which were predominantly mobile homes were undercounted. The housing units in multi-unit structures with 10 to 49 housing units were overcounted. The net undercounts for housing units in multi-unit structures with 50 or more housing units were not significantly different from zero.

The owner and renter status in the tenure variable was an important variable for coverage of persons, but was not significant for coverage of housing units.

5.0 References

- Belin, Thomas R., Diffendal, Gregg J., Mack, Steve, Rubin, Donald B., Schafer, Joseph L., and Zaslavsky, Alan M. (1993) "Hierarchical Logistic Regression Models for Imputation of Unresolved Enumeration Status in Undercount Estimation," *Journal of the American Statistical Association*, 88:423, pp 1149-1159.
- Childers, Danny R. (1992) "The 1990 Housing Unit Coverage Study," *Proceedings of the Section on Survey Methods Research of the American Statistical Association*, pp. 506-511.
- Fay, Robert (1990) "VPLX: Variance Estimates for Complex Samples," *Proceedings of the Section on Survey Methods Research of the American Statistical Association*, pp 266-271.
- Hogan, Howard (1992) "The 1990 Post-Enumeration Survey: An Overview," *The American Statistician*, 46, pp 261-269.
- Hogan, Howard (1993) "The 1990 Post-Enumeration Survey: Operations and Results," *Journal of the American Statistical Association*, 88:423, pp 1047-1060.
- U.S. Bureau of the Census (1993), "Coverage of Housing in the 1990 Decennial Census," HUCS Results Memorandum Number 4, by Danny R. Childers.
- Wolter, Kirk M. (1986) "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, 81, pp 338-346.