

USING HMDA:
EVALUATING THE QUALITY OF
DISCLOSURE DATA FROM 1980 To 1982

BY

ANNE B. SHLAY

FEBRUARY, 1985

WOODSTOCK INSTITUTE 53 W. JACKSON. CHICAGO. IL 60604
(312) 427-8070

(C) 1985 BY WOODSTOCK INSTITUTE

Executive Summary

Any seasoned researcher knows that a study's findings are only as good as the data on which they are based. Great ideas, beautiful language and fancy statistical procedures rarely mask the presence of poor quality, biased and generally faulty information. With ease, one can go after a study by maligning the data base. A study can be disqualified by attacking its most vulnerable and usually incorrigible fatal flaw, the data. One can rewrite a report, redo analyses and rethink ideas. Few projects have the resources to collect new and better data, that is, to begin all over.

Studies using data mandated from the Home Mortgage Disclosure Act (HMDA) have been criticized on the grounds that HMDA data are faulty and inaccurate. This paper shows that such criticisms of research using post-1980 data are unwarranted. The recent availability of computerized HMDA data has allowed a comparison of the quality of these data to those in other well-respected data bases -- the U.S. Census of Population and Housing and the Annual Housing Survey. Analysis of HMDA reports for Chicago over three years showed that the HMDA data aggregated by the Federal Reserve Board are of comparable quality to data aggregated by the U.S. Census Bureau.

Faulty bank reporting and keypunching result in three main types of errors in computerized HMDA data. These errors, which are flagged by the Federal Reserve Board, include incorrect census tracts, inaccurate lending data and missing lending data. Even with these errors, the study showed that 1980 Chicago HMDA data is 89% accurate, 1981 data is 98% accurate and 1982 HMDA data is 99% accurate.

By comparison, the 1980 100 Percent Count U.S. Census of Population and Housing data collection procedures yielded much missing and inconsistent data. To compensate for missing or inconsistent data, the Census Bureau "allocates" or assigns new data values to substitute for missing or faulty information. At least one value was allocated to about 10% of all persons included in the 1980 Census of Population. And the national Annual Housing Survey was found to contain inconsistent values on important items for 1% to 9% of the sample from 1974 to 1979.

Both the Census of Population and Housing and the Annual Housing Survey are respected sources of information which are used to inform public policy and to allocate public monies. The existence and magnitude of errors and inconsistencies within these data bases are similar to those in the HMDA data and suggest that the quality of the HMDA data are soundly within the bounds of public industry standards.

The conclusions of this study suggest that discussions over HMDA should turn from the issue of data quality to those of how to expand this timely data base and to use it more fully.

USING HMDA: EVALUATING THE QUALITY OF
DISCLOSURE DATA FROM 1980 TO 1982

Introduction

Any seasoned researcher knows that a study's findings are only as good as the data on which they are based. Great ideas, beautiful language and fancy statistical procedures rarely mask the presence of poor quality, biased and generally faulty information. With ease, one can go after a study by maligning the data base. A study can be disqualified by attacking its most vulnerable and usually incorrigible fatal flaw, the data. One can rewrite a report, redo analyses and rethink ideas. Few projects have the resources to collect new and better data, that is, to begin all over.

Studies using data mandated from the Home Mortgage Disclosure Act (HMDA) have been victim to the critique that HMDA data are faulty and inaccurate (Canner, 1982a; Canner, 1982b). This paper shows that such criticisms of research using post-1980 HMDA data are unwarranted¹. Based on extensive examination of HMDA data over three years for one major metropolitan region, Chicago, and by comparing the quality of these data to those in other well-respected public use data bases, this paper concludes that aggregated data produced by the Federal Reserve Board provide valid and accurate information on regulated lenders' loan dispositions.

HMDA data consist of commercial banks' and savings and loan associations' disclosure of their residential lending dispositions at the census tract level. Specifically, these lenders disclose by census tract the total number of loans and total dollar amount of lending for the most

common types of residential lending: conventional single-family loans (1-4 units), FHA and VA loans, multi-family loans and home improvement loans. Reported for every Standard Metropolitan Statistical Area (SMSA) in the United States by all regulated institutions with assets of over \$10 million (essentially all SMSA-located regulated lenders), HMDA data permit research which monitors residential credit flows from each lender to the neighborhoods.* Continually provided each year since 1976, HMDA data can be used to assess individual lenders' community reinvestment record over time, to link lending patterns with neighborhood characteristics, to compare residential credit flows among different areas within a metropolitan area and to compare the lending patterns among different metropolitan areas. Coupled with Community Reinvestment Act provisions, HMDA data allow local organizations to deliberately and methodically work to direct more lending in their neighborhoods².

Yet the insistence by some that HMDA data are inaccurate threatens to undermine the credibility of local organizing efforts which use HMDA as well as to garner legislative support for ending the provision of these data altogether. With congressional renewal of HMDA due this year, it is timely to evaluate the quality of HMDA data.

* A Standard Metropolitan Statistical Area is a county containing a central city with a population of 50,000 or more plus contiguous counties which are designated as urban by the U.S. Census Bureau.

This paper evaluates the quality of HMDA data based on experience with using the machine readable HMDA data base produced by the Federal Reserve Board for the years 1980, 1981 and 1982. This evaluation is based on close inspection of the HMDA data for these years for the entire Chicago SMSA. Chicago is the third largest SMSA in the country and contains over 500 regulated lenders (i.e., commercial banks and savings and loans). Because Chicago is the credit marketplace for so many lending institutions and has such an expansive lending territory, Chicago is an optimal test for assessing the accuracy of HMDA data. That is, one might expect a higher rate of HMDA reporting errors and discrepancies in Chicago than in smaller areas with a more concentrated lending industry.

This paper proceeds as follows. First, it explores the types of errors in the Chicago HMDA data base and the extent of these errors from 1980 through 1982. Second, it assesses the quality of these data in light of these inaccuracies and compares these error rates to those accepted in well-respected data bases: the U.S. Annual Housing Survey and the U.S. Census of Population and Housing. Finally, this paper looks toward the future of HMDA data.

HMDA Errors: The Chicago SMSA from 1980-1982

Inaccuracies in HMDA data can occur at two stages of the data compilation process. At stage one, inaccuracies can result with lenders'

failure to report the correct information. At stage two, errors can result from data processing errors, generally key punch errors.

Errors in lenders' reporting may occur in several ways. First, lenders may err in aggregating loans to the census tract level, i.e., they may report lending dispositions for the wrong census tract number. Second, lenders may fail to understand reporting instructions and report lending for the wrong spatial unit, e.g., counties or zip codes. Third, lenders may inaccurately report their lending volume. Finally, lenders may engage in "sloppy" reporting. Sloppy reporting can range from reporting lending for nonexistent census tracts, to mistakes in arithmetic, to reporting lending volumes for the wrong type of loan (e.g., reporting as home improvement loans which are really FHA single-family loans).

Obviously, all types of reporting errors cannot be detected. Nonetheless, a number of checks are possible with the existing data. First, one can assess the validity of reported census tract numbers. Second, one can assess whether loan volumes approximate conventional standards. And finally, machine readable HMDA data can be checked against the hard copy (i.e., paper) reports submitted by each lending institution.

The Federal Reserve's methods for compiling HMDA data into machine readable form facilitates these types of checks. Lenders submit disclosure reports to their respective regulators who generally inspect the data for accuracy and attempt to correct discrepancies directly with each lending institution. The regulators submit these "cleaned" reports to the Federal

Reserve which implements a second cleaning process. Finally, for data discrepancies which cannot be rectified, the Federal Reserve flags each data record for each type of inaccuracy or discrepancy. That is, the Federal Reserve processes (i.e., makes machine readable) any bad data but provides separate indicators to tell the data user whether the data are good or bad and the type of error.

Errors in HMDA data are flagged in two ways. First, data records are flagged if they contain invalid census tract values. The values for this "tract validity flag" variable indicate the type of problem with each invalid tract value (e.g., higher order integer position unmatched, lower order integer position unmatched, tract reported subdivided when it is not, etc.). A simple frequency distribution of this tract validity flag variable indicates the distribution of good versus bad tract values among lenders' reports.

Second, data records are flagged according to the quality of their lending data, i.e., the number of loans made in the census tract and the total dollar amount of lending in the census tract. The lending data flags indicate whether the record contains valid data, whether the record contains a "quality error" and whether the record contains a "validity error."

A quality error indicates that the average loan amount is "out of tolerance," i.e., outside of the range of what is considered by the industry as a normal loan amount. To determine a quality error, the

Federal Reserve constructs "tolerance limits," essentially the loan amount range within which the average loan amount per tract must fall in order to be designated as valid data. Tolerance limits are constructed for each type of lending, e.g., conventional, FHA, home improvement, etc. The Federal Reserve then calculates whether the average loan amount (by loan type) falls within these boundaries (i.e., Total Dollar Amount Lending/Number of Loans = Average Dollar Amount Loan). If the average loan falls outside of these limits, it is flagged as a quality error.

A validity error indicates that the value associated with the number of loans and the value associated with the total dollar amount of lending within a data record do not correspond in a particular way. Essentially, a validity error indicates that either but not both of the number of loans value or the total dollar amount of lending value is zero. This means that the data record may indicate that loans were made in a census tract by a lender but no dollar amount figure is attached to the loans or that the data record shows dollar lending but contains no record of the number of loans.

Again, a simple frequency distribution of each loan flag variable shows the distribution of valid and invalid lending data and indicates the nature of the problem with the data. The distribution of invalid lending data coupled with the distribution of invalid census tracts shows the extent of invalid HMDA data on the machine readable HMDA data files.

Table 1 (pages 21-22) shows the distribution of data errors for the Chicago SMSA for the years 1980, 1981 and 1982. Shown are the total number of lending reports in each reporting year, the number of invalid census tracts and the number of quality and validity errors for each loan type. Shown in parentheses are the total numbers of each type of error as percents of the total number of reports per year.

The data quality in 1980 was poor relative to its 1981 and 1982 counterparts. Approximately 1,500 1980 data records contained invalid tract values. Although 1980 data contained relatively few quality errors (N=361), the data contained over 1,000 validity errors. Clearly, this year of HMDA reporting represented a year of confusion or sloppiness for lenders and/or regulators who reported this faulty information. Yet, by 1981 this confusion/sloppiness was considerably reduced. 1981 Chicago HMDA data contained only 194 invalid census tract values, only 200 quality errors and was devoid of validity errors. The 1982 Chicago HMDA data was "cleaner" than its 1981 counterpart, although the change from 1981 to 1982 was not dramatic.

Nonetheless, the overall change in data quality over this three-year period shows massive improvement in HMDA reporting and data processing procedures. From 1980 to 1982, the number of invalid census tracts was reduced by 92%. The number of quality errors was reduced by 57%. The number of validity errors was reduced almost entirely. 1981 and 1982 HMDA data for the Chicago SMSA are fairly accurate representations of the lending activity by regulated lenders in this SMSA.

Yet as discussed below, the relative amount of invalid data even in 1980 may not be high in terms of conventional data standards. Although approximately 1,500 **invalid census tracts** are contained in the 1980 Chicago data base, this represents only 5% of the total number of reports (N=27,291). The total number of 1980 **quality errors** represent only 1% of the total reports. The total number of 1980 **validity errors** represent only 4% of the total reports. Clearly, this data base does not represent with absolute precision regulated lenders' activity in the Chicago SMSA. Yet, approximately 89% of a data base representing 27,291 data points contained "clean" and presumably valid information.

What is Dirty Data: HMDA Comparisons with Other Data Bases

The utility of any data set is based on its ability to be used to answer questions. Any assessment of a data set should be based on the level of accuracy and precision required to answer particular questions. That is, some questions may require data of greater precision than others. Hence, what may be deemed dirty or invalid data from the vantage of one research purpose may be clean and perfectly useful data from the vantage of a different research purpose. Determining data quality is a judgment based on subjective standards which arise out of a research situation.

Applied to disclosure data, the question becomes, what is the requisite level of precision in order to address lenders' individual performances, the distribution of lending across metropolitan areas,

lending discrimination, disinvestment and lending inequities. For example, more precision may be required in order to address the performance of single lenders where the data are specifically targeted to show where an institution is or is not making loans. Less precision may be needed in assessing lending distributions across a locality for an aggregation of lenders. Of course, in assessing whether the data are adequate to answer any question one must also decide if by using partially biased data which can be used to improve public welfare (e.g., to introduce reinvestment issues to particular lenders), may be better than no data at all.

These issues face users of any data set regardless of whether the data are collected by the researcher or by some other agent. This part of the paper examines the quality standards imposed on two public use data bases and compares these to the error rates found in the Chicago HMDA data bases.

The first data base examined is the 1980 U.S. Census of Housing and Population 100 Percent Count. These data were enumerated for the total composite of housing and population within U.S. boundaries. Since HMDA data are also "100 Percent Counts," i.e., represent the total composite of regulated lenders' home mortgage activities within every SMSA, assessing HMDA data quality relative to the Census 100 Percent Count is appropriate; both are only subject to non-sampling errors. The second data set examined is the U.S. Annual Housing Survey. The Annual Housing Survey represents a sample of U.S. housing. Data for these housing units have been collected each year since 1974. The longitudinal structure of the data set (i.e., data for the same housing units over time) facilitates tracking data

discrepancies and inconsistencies (i.e., errors) over time and therefore, allow more ease in discerning valid from invalid data.

To be sure, employing the 1980 Census of Population and Housing and the Annual Housing Survey to determine public data quality standards is not because either of these data sets represent the "best" data available. Nonetheless, both of these data sets represent government collected data which are applied to major public policy and benefit distribution issues facing federal, state and local agencies as well as corporations and neighborhood organizations. And they are collected and managed under the auspices of a large federal agency, the U.S. Bureau of the Census, one with vast experience in applying rigorous standards to enormous volumes of information. In this light, HMDA data quality is compared to these public industry standards.

The U.S. Census of Population and Housing

The 1980 Census of Population and Housing enumerated basic demographic, socio-economic and housing information on over 226.5 million people and on over 86.5 million housing units. Obviously, within such a large scale data collection process, errors occur.

The Census Bureau routinely and systematically evaluates data comprehensiveness and quality in terms of a number of criteria. First, the bureau determines its success in enumerating all housing units and individuals (i.e., did it count everyone and every house?). Second, it

determines the completeness of information (i.e., were all survey questions completed?). Third, it determines data inconsistencies among respondents (e.g., an unmarried man with a wife present or an owner-occupied unit with a monthly rent). Finally, it determines data processing errors (i.e., transmission of survey items into machine readable form).

Rather than ignore missed units or respondents or delete faulty, incomplete or inconsistent data, the Census essentially "corrects" the data from available information concerning other respondents or housing units. This process is generically termed "allocation" or one in which data values are allocated or assigned to missing or faulty data entries. Allocated data represent a reasonable "best guess" about true data values. For example, a rent level would be assigned to a unit with a missing rent level based on the unit's comparability with another unit which has the necessary information. A person's income would be allocated based on the income of another person with similar characteristics (e.g., age, education, etc.). The method for allocating data does not disturb the distribution of population parameters (i.e., the mean and standard deviation) so that the overall distribution of values is constant before and after the data are allocated. **Nonetheless, allocated data values for individual members of the population may or may not be correct.**

In addition to allocating single data values or a series of data values to a person or housing unit, complete sets of data values are assigned when no information is available. Essentially, the

characteristics of the previously counted person or housing unit are substituted for the missing observation.

The generation of information where data are missing is not an unusual procedure when the appropriate methods are employed (Babbie, 1983)³. Data sets which contain allocated data are not necessarily poor data bases. Yet by including allocated data, these data bases contain a proportion of information with unknown reliability. Allocated data are imperfect data; they could be incorrect. For the purposes here, the extent of allocated census data represent the tolerance levels by the census for imperfection within their public use data sets. Table 2 (pages 23-24) shows the extent of housing data allocation.

For all housing within the Chicago SMSA and the United States respectively, Table 2 shows the number and percent of allocations for selected housing characteristics. Each percent is based on the total number of housing units for which each selected housing characteristic is collected. For example, the 1.3% associated with "lacking complete plumbing facilities" means that the data values for the variable "lacking complete plumbing facilities" was missing and thus allocated for 1.3% of all year-round housing units in the Chicago SMSA as opposed to occupied or vacant units.

This table shows a wide range in the extent of allocation among these housing characteristics. Only 1.7% of all Chicago SMSA year-round housing units and only 2.3% for the United States as a whole contained allocated

data for the variable "number of units at address." Yet, approximately one-third of all vacant units for both Chicago and the country contained allocated data for the variable "rent asked."

Nonetheless, when even the smallest percentages are translated into actual numbers of housing units for which data were allocated, the numbers are very large. That 1.7% of all Chicago SMSA year-round housing contained allocated data for the variable "units at address" means that for about 45,000 Chicago housing units, data values were generated for this variable. Since the average 1980 Chicago SMSA census tract actually contained a mean of 1,738 housing units, the 44,806 allocated data points represent the entire housing stock of approximately 26 "average" census tracts.

For those housing characteristics which had higher allocation rates, the number of housing units for which data values were allocated may appear alarmingly high. For 11,764 vacant housing units in the Chicago SMSA and for 504,252 in the U.S., whether these units were boarded up or not was allocated based on the status of adjacent housing. For approximately 20,000 Chicago vacant housing units and for over 1 million U.S. vacant housing units, the asking price was allocated. Although the U.S. Census of Housing may represent the best and most comprehensive housing data base for the country, clearly it is not perfect.

Table 3 (page 25) shows similar information for selected population characteristics. Included in this table is the "substitution rate" (i.e., the number of percent of all persons for which **entire data records were**

generated) and the "overall allocation rate" (i.e., the number and percent of all persons for which one or more data values were allocated).

For the Chicago SMSA, data for over 200,000 persons was completely created from the data associated with other persons. Over ten percent of all persons in the Chicago SMSA and in the country had at least one data value allocated representing over 700,000 Chicago SMSA residents and over 3,000,000 U.S. residents. Since the average 1980 Chicago SMSA census tract contained a mean population of 4,686 persons, the Chicago substitution rate represents the entire population of approximately 45 "average" census tracts; its allocation rates represent the population of approximately 158 census tracts.

The point here is not that census data are poor or that the U.S. Census Bureau is not effectively carrying out its mandate. On the contrary, these allocation rates may be interpreted as an extremely able record in data collection and processing procedures given the extent of the task. Nonetheless, the extent of allocation shows that this trusted data base is not error free.

The Annual Housing Survey

The longitudinally-linked national Annual Housing Survey (AHS) contains housing and household data from a yearly survey of approximately 75,000 U.S. housing units and their occupants since 1974. Because the same housing units are included each year, the data permit more direct analyses

of housing change and residential turnover. The data have been used extensively for informing housing policy decisions.

Although the AHS sample design is longitudinal, the census' data processing procedures have treated the data cross-sectionally. In other words, the allocation of data estimates to observations where information is missing is based on the characteristics of the sample within each specific year and is not based on information from a previous year. For example, an observation which has missing information pertaining to whether the house has a basement in 1978 would be assigned this data value based on the values of other housing units in 1978 independent of whether this house was shown to have a basement in 1977. In short, housing units over time could vary on major and presumably non-changing structural characteristics simply on the basis of the way missing data are treated.

The construction of the longitudinally linked AHS data base by the Annual Housing Survey Project at Columbia University permitted systematic examination of the data over time. A report issued in 1982 by the co-director of this project showed that serious problems existed in the data base from data inconsistencies on a year-to-year basis (Beveridge, 1982). For consecutive pairs of years, the type and size of housing units fluctuated over time for an average of 5% of the sample. The number of rooms per unit fluctuated for an average of over 9% of the sample. Whether a unit was in or out of a public housing project fluctuated over time for almost one quarter of the sample. In short, basic yearly information on

the structural characteristics of housing units changed each year for a large proportion of the sample.

In addition, household characteristics fluctuated over time when the household head was reported to be the same individual for consecutive years. The report shows that respondents' race or sex changed over time although the respondent per se was constant over time. Race changes occurred for an average of approximately 1% of the sample and sex changes occurred for approximately 2.5% of the sample.

The AHS data inconsistency rate is too high to be explained by the methods used to allocate data alone. Indeed, the source of AHS data inconsistencies remains a mystery (Beveridge, 1982). To be sure, by showing these data inconsistencies in this large and important data base, the report's author was not solely concerned with the quality of the Annual Housing Survey. He suggests that data inconsistencies in the AHS may appear more extreme than in other large scale data bases because the AHS longitudinal structure permits ease in detecting inconsistencies. In short, other large data bases may be plagued with large amounts of unknown data errors⁴.

The Future of HMDA

HMDA data, the U.S. Census and the Annual Housing Survey are all large-scale surveys. HMDA is a different type of survey in that the survey instrument is not a questionnaire but is a standard form supplied to

lenders to systematically report their residential lending dispositions. To conform to the Home Mortgage Disclosure Act, lenders must simply aggregate their lending to the census tract level. The task is clear.

Nonetheless, one would expect that household survey instruments such as the AHS and the U.S. Census provide even clearer instructions for enumerators and respondents. Survey questions are always subjected to rigorous pre-testing in order to remove any ambiguity. Errors occur despite infusions of time, money and caution. Indeed, this paper has shown that systematic efforts at compiling the best and most accurate data produce results that are far from perfect. Yet the U.S. Census and the Annual Housing Survey are used to inform major policy decisions, to distribute public monies and to supply information. Recent battles over the accuracy of SMSA population counts are taking place because millions of public dollars can be lost to a municipality if population size decreases. From the public's perspective, the costs of census data errors are large.

To be sure, if important issues and decisions did not rely on the knowledge data provides, its accuracy would be irrelevant. But caution must be exercised in making facile judgments over what are high or low quality data. As has been shown, even the U.S. Census Bureau with a lengthy history of developing and testing methods faces rectifying errors and missing data for each and every census taking. And like HMDA, the sophistication of methods and the accuracy of the data have increased over time.

Three issues are important here. First, public use data bases may not contain low accuracy rates but rather, our standards for quality data may be too high. Expending more money to create a 99% accurate data base over one which is 95% accurate may not be cost effective. Will we know substantially more about lending discrimination with an additional 800 lending reports when we already have the information provided from 20,000 reports for a single SMSA in a single year? Will our insights into neighborhood disinvestment be biased (i.e., wrong) because we lack a small proportion of correct data? It is doubtful. Indeed, current HMDA accuracy rates may be perfect for the job. Given the standards applied to other major data bases, one might even argue that the accuracy rates exceed what is necessary.

Second, the judgment that HMDA data are quality data and have improved over time is one based on technical criteria which cloaks the politics of the issue. Independent of the facts concerning these data, for some, HMDA data may never be "quality." For lenders who dislike disclosing their institution's decisions, for regulators who do not want to process the information, and for those who choose to argue that the market works and works fairly despite evidence to the contrary, HMDA data will continually be subjected to critique. With the caveat that HMDA data are bad data, analyses can be easily ignored. The argument that HMDA data are poor quality is easier to digest than the argument that lenders do not want to be publically accountable for their decisions. As mayors scream "the Census is wrong" because they do not want to bear the fiscal consequences of their shrinking cities, others may shout that HMDA is wrong because they

do not want to be held responsible for urban decay and credit inequities. Disqualifying HMDA channels political discussions away from the public costs of private disinvestment and the responsibilities of privileged private lending institutions.

Finally, HMDA is a young data base. Disclosure was mandated by the federal government less than ten years ago. Given the confusion that surrounded the initial implementation of HMDA, one should not expect that the early years of HMDA would provide the best information. Yet with legislative revisions in 1980 which mandated standard reporting forms, central depositories, and the processing of all HMDA data into machine readable form along with lenders learning how to systematize their methods of reporting, HMDA reports became reliable sources of information. Legislative affirmation, government purpose, advanced lender learning curves and the hard work of local regulators and the Federal Reserve Board have worked to produce a fine data base.

With the recent availability of HMDA data in machine readable form, people can begin to comprehensively examine metropolitan lending patterns. Knowledge about private investment decisions is critical in an era of declining public revenues for cities. Today, public revenues will be even less likely to fill in for absent private capital. Now is not the time to think about dismantling a data collection and aggregation process which took time and resources to implement and which currently provides quality and necessary information. More appropriately, given the success of all parties in conforming to HMDA legislation, legislators might begin to

construct more inclusive disclosure requirements and mandate the reporting of commercial loans as well as disclosure by non-regulated lenders. The questions should turn quickly from "is HMDA any good" to "what other kinds of data can complete a timely and reliable source of public information."

TABLE 1

Errors in the Chicago SMSA HMDA Data Base From 1980-1982

	<u>1980</u>	<u>1981</u>	<u>1982</u>
<u>Total Number of Reports</u>	27,291	23,247	19,808
<u>Invalid Census Tracts</u>			
Number	1,496	194	115
(% of total reports)	(5.48%)	(0.83%)	(0.58%)
<u>Quality Errors</u>			
Conventional, Single-family			
Number	160	77	124
(% of total reports)	(0.59%)	(0.33%)	(0.63%)
FHA			
Number	50	71	7
(% of total reports)	(0.18%)	(0.31%)	(0.04%)
Home Improvement			
Number	47	30	17
(% of total reports)	(0.17%)	(0.13%)	(0.09%)
Multifamily			
Number	79	15	1
(% of total reports)	(0.2%)	(0.06%)	(0.01%)
Non-Occupant			
Number	25	7	5
(% of total reports)	(0.09%)	(0.03%)	(0.03%)
Total Quality Errors			
Number	361	200	154
(% of total reports)	(1.32%)	(0.86%)	(0.78%)
<u>Validity Errors</u>			
Conventional, Single family			
Number	655	0	1
(% of total reports)	(2.40%)		
FHA			
Number	89	0	0
(% of total reports)	(0.33%)		
Home Improvement			
Number	376	0	0
(% of total reports)	(1.38%)		

TABLE 1

	<u>1980</u>	<u>1981</u>	<u>1982</u>
<u>Validity Errors</u> , cont.			
Multifamily			
Number	44	0	0
(% of total reports)	(0.16%)		
Non-Occupant			
Number	27	0	0
(% of total reports)	(0.10%)		
Total Validity Errors			
Number	1,191	0	1
(% of total reports)	(4.36%)		

TABLE 2

U.S. Census of Housing Allocation Rates for Selected Housing
Characteristics for the Chicago SMSA and the United States, 1980

	<u>Chicago SMSA</u>	<u>United States</u>
<u>Year Round Housing Units</u>		
Total Number	2,635,635	86,692,823
Lacking Complete Plumbing Facilities		
Number Allocated	34,263	866,928
Percent	1.3%	1.0%
Number of Units at Address		
Number Allocated	44,806	1,993,934
Percent	1.7%	2.3%
Condominium Status		
Number Allocated	79,069	3,034,248
Percent	3.0%	3.5%
Number of Rooms		
Number Allocated	55,348	1,473,778
Percent	2.1%	1.7%
<u>Occupied Housing Units</u>		
Total Number	2,486,724	80,389,673
Tenure		
Number Allocated	57,195	1,607,793
Percent	2.3%	2.0%
Value - Specified Owner		
Number Allocated	67,142	3,617,535
Percent	2.7%	4.5%
Value - Condominium		
Number Allocated	77,088	2,894,028
Percent	3.1%	3.6%
Contract Rent		
Number Allocated	84,549	2,331,300
Percent	3.4%	2.9%
<u>Vacant Housing Units</u>		
Total Number	148,911	6,303,150
Vacancy Status		
Number Allocated	7,892	359,279
Percent	5.3%	5.7%

TABLE 2

	<u>Chicago SMSA</u>	<u>United States</u>
<u>Vacant Housing Units, cont.</u>		
Duration of Vacancy		
Number Allocated	21,890	1,027,413
Percent	14.7%	16.3%
Boarded Up		
Number Allocated	11,764	504,252
Percent	7.9%	8.0%
Price Asked		
Number Allocated	19,954	1,040,019
Percent	13.4%	16.5%
Rent Asked		
Number Allocated	48,247	2,294,346
Percent	32.4%	36.4%

Source: U.S. Department of Commerce, Bureau of the Census, General Housing Characteristics, United States Summary, Table A-2, 1980.

TABLE 3

The Census of Population Allocation and Substitution Rates for Selected
Population Characteristics for the Chicago SMSA and the United States, 1980.

	<u>Chicago SMSA</u>	<u>United States</u>
<u>Total Persons</u>		
Number	7,103,634	226,545,845
<u>Substitution Rate</u>		
Number	213,109	3,398,187
Percent	3.0%	1.5%
<u>Overall Allocation Rate</u>		
Number	738,778	23,107,676
Percent	10.4%	10.2%
<u>Allocations - Specific Indicators</u>		
Relationship		
Number	170,487	4,757,463
Percent	2.4%	2.1%
Sex		
Number	56,829	1,812,366
Percent	0.8%	0.8%
Age		
Number	213,109	6,569,829
Percent	3.0%	2.9%
Race		
Number	113,658	3,398,187
Percent	1.6%	1.5%
Ethnic Origin		
Number	284,145	9,514,925
Percent	4.0%	4.2%

Source: U.S. Department of Commerce, Bureau of the Census, General
Housing Characteristics, United States Summary, Table B-4, 1980.

Footnotes

1. To be sure, Canner reports of a 1979 study which showed the poor quality of a sample of 43 reporters' HMDA statements within three SMSA's in 1977. This research, however, was used to develop methods to improve the accuracy of HMDA, and led to the Federal Reserve Board taking steps to provide clearer instructions to lenders as well as to mandate a standard reporting format (Canner, 1982b).

2. The Community Reinvestment Act of 1977 permits individuals and local community organizations to challenge lender mergers, relocations, acquisitions and federal charters if lenders can be shown to be unresponsive to community credit needs. HMDA data provide the necessary information to assess where each lender's housing loans are being made.

3. A major text book in research methods suggests several appropriate techniques for treating missing data besides eliminating data points altogether (Babbie, 1983: pp. 377-378). The choice of method can be determined by the research task as long as the generation of new data values does not bias research findings in any particular direction.

4. My personal experience with using the longitudinally linked Annual Housing Survey confirmed that the data was fraught with inconsistencies. Indeed, one important variable, household income, had to be dropped out of the analysis altogether because the data were so heavily allocated. Household income measures varied so widely over time within constant households that one could not reliably discern actual household income changes from those which were artifacts of data allocation methods.

References

Babbie, Earl. The Practice of Social Research. Belmont, California:

Wadsworth Publishing Company, 1983.

Beveridge, Andrew A. "Reliability in Large Scale Household Surveys: The Case of the Annual Housing Survey." Paper presented at the Annual Meeting of the Association of Public Data Users, Washington, D.C., October, 1982.

Canner, Glenn B. "Redlining Research and Federal Legislative Response."

Federal Reserve Bulletin, Volume 68 (October), 1982a.

Canner, Glenn B. "Home Mortgage Disclosure Act Aggregation: Benefits,

Accuracy and Costs." Journal of Retail Banking, Volume 7,

Number 4: 37-46, 1982b.

U.S. Bureau of the Census. General Housing Characteristics, United States

Summary. Washington, D.C.: U.S. Department of Commerce, 1980.

U.S. Bureau of the Census. General Population Characteristics, United

States Summary. Washington, D.C.: U.S. Department of Commerce, 1980.