

---

# Administrative Records Use in the 1995 Census Test\*

Joseph J. Knott, Bureau of the Census

---

The 2000 Census Research and Development Program is developing the information needed to decide the utility of administrative records usage in the 2000 census. The 1995 Census Test is the culmination of this research and development program. This paper describes the uses of administrative records to be tested in the 1995 Census Test and how these uses will be evaluated. The administrative records to be sought will be listed, along with the rationale for each and any issues involved in the acquisition of each. The "Notice" as required by the Privacy Act will be discussed in terms of its implications for the 1995 Census Test, along with privacy and informed consent issues. Enumeration rules, as they relate to the use of administrative records in a census, also will be discussed.

## ■ Census Test -- Background and Description

In 1991, the Bureau of the Census established the Year 2000 Research and Development (R&D) Program to identify new approaches that would make fundamental improvements in the 2000 census. These efforts were in response to changes in society and demands for a less costly and more accurate census.

Changes in society, as reflected in the lower than anticipated mail response rate for the 1990 census, were a major motivator for developing fundamentally better census taking procedures. Because of these changes, some of the 1990 census methods will *not* provide a more accurate census at a reasonable

---

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

cost in 2000. We need new and improved methods to account for changing household structures, decreasing public cooperation, and the increasing diversity of our population.

The costs of the decennial census have increased every decade, resulting in the cost of the 1990 census being \$2.6 billion. It is estimated that if we were to take the census the same way in 2000, the cost of designing, planning, preparing, and implementing it would be more than \$4.8 billion.

The accuracy of the 1990 census count was judged to be not adequate by many, even though it included almost 99 percent of the population. Of particular concern was the fact that minority population subgroups were missed at higher rates than non-minority subgroups. This *differential undercount* affects the major purposes of the decennial census -- equitable apportionment of Congressional seats among the several states; fair redistricting within states; and funding allocations based on census population data.

Therefore, the R&D Program was assigned the task of developing new and improved procedures that would reduce the differential undercount and contain the cost of the census. The changes being tested in the 1995 Census Test range from more extensive use of the U.S. Postal Service to the extensive use of sampling and estimation. One of the changes with the greatest potential for future application is the use of administrative records to reduce the size of the population undercount and to help control costs [1]. This paper will concentrate on that aspect of the R&D program.

## ■ Counting, Assignment, and Estimation in the Census

There are three ways to be counted or enumerated in the decennial census -- counting, assignment from administrative records, and estimation. To un-

derstand the administrative records research in the 1995 Census Test, it is important to appreciate these concepts.

*Counting* uses the full array of direct contact techniques to reach respondents -- by mail, by personal visit, by telephone, or by other means. Counting also includes data obtained by proxy for another household, housing unit, or person. Historically, the Census Bureau has added people to the census by using information from administrative records as an indicator and then verifying this information. In previous censuses, this approach also was considered counting and it was done only on a limited scale.

By *assignment* we mean using indirect evidence from administrative records to "assign" previously uncounted people to a specific geographic location or address, without further verification. It is the hypothesis of the administrative records research that administrative records contain people that are missed in traditional census counting procedures. There are several ways of using administrative records to improve decennial census accuracy. One way is to add a person to the census based on her or his existence on one or more administrative records. The various possible "rules" for using administrative records this way will be explored in the 1995 Census Test. A second way is to use administrative records as part of the procedure for measuring the accuracy of the census count, as part of methodology involved in a statistical correction process. In the 1995 Census Test, the accuracy of the count will be measured using an independent reinterview.

Household reinterviewing is subject to many of the same deficiencies as counting techniques, and administrative records have the potential to reduce the negative effects of these deficiencies. This aspect of using administrative records will be evaluated by a post-reinterviewer match of the people in the administrative records set to the people identified as "uncounted" by reconciliation of the census and the reinterviewer. Those people identified as "real but not counted" using this experimental administrative records procedure will *not* be included as part of the statistical correction process used for the 1995 Census Test.

Third, summary information from administrative records can be used to provide more current information about localized demographic and housing conditions than one can obtain by relying on small area data from the previous census. Having updated information for small geographic areas will help the Census Bureau better target the areas in which it can effectively use special enumeration methods, such as blitz enumeration, alternative language forms, and so forth.

By *estimation* we mean the application of an array of statistical techniques to account for people or housing units not directly counted or assigned. Estimation encompasses those processes that include performing all or part of the enumeration at only a sample of nonresponding households and then using statistical techniques to estimate the populations totals and characteristics for all those households; performing some appropriate coverage measurement procedure and integrating the resulting estimates into the final census result; and other types of modelling or imputation.

The two aspects of statistical sampling and estimation being developed in the 1995 Census Test are *sampling of Nonresponse Households to develop data for all nonresponse households* and *Integrated Coverage Measurement* as a way to *measure and statistically correct the population "counts" as an integral part of the census taking process*.

The 1995 Census Test will use a Census Day of March 4, 1995, in three urban sites and one rural site. The rural site consists of six parishes in the State of Louisiana. The urban sites are Oakland, CA; New Haven, CT; and Paterson, NJ [2].

## ■ Administrative Records Testing Plan

Not everyone is counted in a census. As estimated by the Post Enumeration Survey, the Census Bureau did not count 1.6 percent of the population in the 1990 census. The people the Census Bureau believes it missed varied by age, sex, race, Hispanic origin, tenure, and other classifications. For example, the estimated undercount of the non-Hispanic white population was 1.2 percent, while the esti-

mated undercount of the African-American population was 4.4 percent. For all individuals aged 0 to 17 years, the undercount is estimated at 3.2 percent, while individuals aged 30 to 49 years were undercounted by an estimated 1.4 percent.

About one-third of the people the Census Bureau believes it missed in the 1990 census were in housing units that the Census Bureau did not enumerate for various reasons. The remaining two-thirds of the people the Census Bureau believes it missed were at an address that was part of the 1990 census address list, but were not counted for various reasons. The premise of the administrative records research is that people not responding to the census are identified on one or more lists maintained by government or other sources. For example, most adults have a drivers license and most pay Federal, state, or local taxes, so their names will appear on one of those record sets. Other types of record sets can identify many low income rental housing units. The first research objective is to determine which of the several administrative records sets the Census Bureau can use to identify and add people otherwise missed in the enumeration process. From this information the Census Bureau can determine whether this methodology would provide another effective way to reduce the magnitude of the differential undercount.

Other research objectives of the administrative records program for the 1995 Census Test are to evaluate the amount of duplication that exists among the several sources and to determine *which* sources are most useful for identifying missed people and conducting a more accurate census at the least possible cost. To do this in the 1995 Census Test, the Census Bureau will seek files at the Federal, state, and local levels and evaluate the processing difficulty, utility, and cost effectiveness of each.

For the 1995 Census Tests, the lists the Census Bureau plans to request appear in Figure 1. Part of the research is to determine which lists the Census Bureau can obtain for statistical purposes. We have *not* yet obtained permission from all these agencies to use their lists. Discussion with the responsible agencies began in March 1994, with the goal of obtaining needed permissions by the end of 1994. The

files being sought all have the potential for containing people or addresses that would not otherwise be included in a census of population and housing.

The TRACS file being developed by the U.S. Department of Housing and Urban Development (HUD) will contain the addresses of many rental-subsidized housing units, along with some demographic information. The Medicare files provide excellent coverage of the older population. The NUMIDENT file does not contain current information, but can be used to add demographic information to other administrative records via a Social Security Number (SSN) match.

The IRS data has broad coverage. The request for an extract from the IRS Informational Tax Return file is a first time request by the Census Bureau for these data. Research by IRS staff suggests that the combination of individual and information returns will provide excellent coverage [3].

About one in ten people in the United States participate in the Food Stamp program at any given time. The coverage is disproportionately larger for minority groups and people residing in rental units. The WIC, AFDC, and Medicaid programs also tend to focus on specific segments of the population, and some proportion of the participants in these programs do not receive food stamps. The expectation is that by acquiring all these files, the accumulative coverage will be improved somewhat.

Drivers license records provide good coverage of the late teen and adult populations, but most states now issue drivers licenses for a multi-year period. As a consequence, the currentness and accuracy of the address on these files may not be good. This is an aspect of the drivers license research.

Public school enrollment records also are being tested in the 1995 Census Test. Because the schools collect this information yearly and use it for assigning children to specific schools, the assumption is that the address information on these records is current. Coverage of school-age children is a population group that school records may be able to address directly. For the same reason, the Census Bu-

**Figure 1: Administrative Records Sets the Census Bureau Will Try to Acquire  
in Conjunction with the 1995 Census Test**

Federal	State	Local
TRACS (Tenant Rental Assistance Certification System)	Food Stamps*	School enrollment
	WIC* (Women, Infants, and Children) Medicare	Tax assessment
NUMIDENT (Social Security Number System)	AFDC* (Aid to Families with Dependent Children)	Telephone subscribers**
	Medicaid	Voter registration
1993 and 1994 IMF (Individual Tax Returns)	Drivers license	Utility**
		Rental properties**
1993 and 1994 IRMF (Informational Tax Returns)		Parolee/probationer***
		Head Start

\*State level maintenance of the file  
 \*\*Private company maintenance of the file  
 \*\*\*Local level maintenance of the file

reau also will attempt to get data from a few Head Start "contractors."

The Census Bureau used parolee and probationer files to improve the coverage of this population in the 1990 census and will study this again as part of the 1995 Census Test research. Tax assessment, telephone subscriber, and utility company records also will be sought as potentially useful sources of housing unit addresses for this test, and the people listed on these records also will be part of the person matching process. Census Bureau staff also are trying to locate local or commercial files of rental properties that can help improve "within structure" unit designation.

**■ Operational Access Procedures and Issues**

The administrative records being sought by the Census Bureau are decentralized. Although some files, such as the Food Stamp records, are maintained at the state level, Federal agency permission is required to gain access to these data. Consequently, considerable work is involved in obtaining permission and acquiring the records needed for the planned research. In the 1995 Census Test, the Census Bureau staff seeking the records are not applying a "practicality" criteria for obtaining and using a particular record set in a full census environment; the goal is to assess the value of each record set. Simi-

larly, the Census Bureau is not considering as a criterion the uniformity of coverage a particular record set might provide on a nationwide basis. This was a conscious decision. We do not, in fact, know how hard it would be to acquire a full set of locally maintained records for the 2000 census, nor do we know about the uniformity of content or the maintenance and coverage standards applied for various lists across state and local agencies. The goal for 1995 is to assess the utility and value of a particular file; the issues of practicality and uniformity will be considered in future testing.

Operationally, the identification of, request for, and negotiations associated with obtaining the desired administrative records files is being done by six Census Bureau employees working part time on this project. Each of the four 1995 test sites [2] has been assigned to one person, to provide continuity in dealing with state and local officials. In announcing the selection of the test sites and in subsequent discussions with the highest elected official in each site, Census Bureau staff requested that each site appoint a local coordinator to ensure a focus in requests for local files. This has been helpful, but, for the 1995 test, the Census Bureau staff directly initiated contact with local agencies to request files in advance of the sites naming a coordinator. In future tests, the Census Bureau will try to determine the extent to which it could ask the state and local governments to acquire and provide the desired files without such extensive staff effort.

## ■ Processing

The processing of the administrative records obtained for the 1995 Census Test is being done mainly on a cluster of workstations linked by a Local Area Network (LAN) to a restricted number of terminals. There are two uses of the administrative records data that are being tested that are not discussed in this paper; they are being accomplished on other Census Bureau computer systems.

- The *planning database* is a collection of data at the census tract, block group, and block level that will be used to guide the choice of

areas in which to apply particular special enumeration methods. Tabulated administrative records data will be part of this database. For example, the number of food stamp recipients in a block may be a proxy for current levels of poverty.

- The second involves a post census evaluation (using administrative records such as assessment records that are being acquired for this analysis) to find other sources for addresses not identified by the *Master Address File (MAF) updating* procedures.

Importing the various administrative records generally is being done using whatever basis is acceptable to the supplier. Although the Census Bureau has a set of standards that define the preferred media, file, and record formats, the staff has chosen to be flexible in the 1995 test. For this research, the Census Bureau will try to get the data in one of the standard formats, but if this is a problem for a local agency, the Bureau's programming staff has chosen to be very accommodating. (This has even gone to the point of taking a paper listing and keying the data.) In a larger program, this level of flexibility would be impractical.

The basic processing required is to reformat the administrative records data into a standard format. To the extent the data have an SSN available, the data are unduplicated and a master person record is created. This record is updated or created with the merger and unduplication of information from all files received. There is a separate master administrative records file for each of the four 1995 Census Test sites.

Although most of the administrative records do, in fact, have an SSN, some do not or the providing agency has chosen not to include them. For example, the California Department of Motor Vehicles has SSNs in its drivers license data set, but will not provide them.

As multiple records for the same SSN are merged, a set of rules are applied that determine the primacy

of each data item. For example, the address from a Food Stamp program participant would be preferred over the address from a drivers license record, because it is thought to be more current. Although each data item on the master person record has only one value, the software stores a "walk back" record number for each matching record on the master record, so that all possible sources of a data item can be analyzed, if necessary. Further, a source flag is maintained by item, so that the source of each data item on the master person record can be ascertained easily.

When the person processing is completed, the SSN-sorted file will be inverted to an address-based file and merged with records that have only an address to use as a match key -- records such as those on the California drivers license file. This address-based file then will be matched to the MAF for analysis -- the MAF being the list of all census addresses in the test area.

The final analytical system includes three datasets: the set of administrative records assigned to an address, the MAF, and the 1995 Census Test questionnaire data. It should be noted that the 1995 Census Test questionnaire data set includes the full name reported for each household member, as well as the birth date (month, day, and year), gender, race, and Hispanic origin.

## ■ Match Key Research

The matching keys and the performance of each is an important area of research.

The five matching keys are:

- **Address** -- This is a primary match key, in that it allows entry into the Census Bureau's other files using its address matching and/or geocoding systems.
- **SSN** -- The SSN is an excellent key and can be used to link people from different administrative records sources and to unduplicate administrative records lists; SSNs will not be collected during the census, however, and

this will not be useful as a key for matching responses or questionnaires.

- **Date-of-Birth, Gender, and Name** -- These items provide a powerful matching key in a person record data base. In the Microform sample of the Simplified Questionnaire Test (SQT), the Social Security Administration provided NUMIDENT data, including name, date-of-birth, SSN validation code, race (or Hispanic Origin), and sex. With these data, the Census Bureau has an unique dataset containing both administrative records data and survey-collected data for the same set of people. The research is to determine the degree of confidence that can be attributed to a match using various items individually and as groups. The conclusions reached from this research will be used to help develop and improve future matching rules. For example, date-of-birth is a good match key, but how good is it when compared with SSN?
- **Household Composition** -- This is a "new" matching key, in that it tries to use commonly available data, such as sex and age, as match keys by grouping household members together to form a new key. There are a number of issues related to this approach and the research on this key will need to be continued. This key will not be used in the 1995 Census Test.
- **Telephone Number** -- The use of telephone number as a match key appears to have promise. Telephone number will not be used as a match key in the 1995 Census Test.

## ■ Enumeration Rules

To be useful in supplementing the decennial census enumeration process, administrative records must be available at the address level while the field operations are being done. In order for this to happen, administrative records must be integrated into the "MAF system" shortly before the MAF is made available for field operations.

Throughout 1994 and early 1995 the selected administrative records will be collected and processed into the administrative records database. At some point, as close to Census Day [4] as possible, a set of *assignment rules* will be run on the administrative records data and specific administrative records and information will be linked to the MAF at the address level. Basically, the information will be of three types:

- **Whole Household Assignment** -- A set of administrative records will exist that may be used, in the absence of a completed census form, for the enumeration of an entire household. The criteria for accepting a set of administrative records as a household is, "If one were to enumerate the housing unit by traditional means (mail, personal visit, or telephone), would the information collected be the same as what one could derive using the administrative records set?" In other words, were the "administrative records" people, and only those people, residing at that address on Census Day? Census Bureau staff postulate that administrative records will perform as well as the traditional modes of follow-up, and better than allocation or weighting, in answering this question. To develop the specific assignment rules, the following information will be considered:
  - **Address currency** -- In general, the Census Bureau will use what it considers to be the most current address from all available administrative records sources.
  - **Mover status** -- If the administrative record has an indication of a move in the past 6 months, the address will not be used as the preferred address.
  - **Does a whole household administrative record exist?** -- Record sets, such as those for Food Stamps, AFDC, or TRACS, generally represent whole households. For example, the TRACS file has information about each person in a subsidized

rental unit. An issue here is whether the data set must contain race/ethnic data to be adequate? These data most often are absent from administrative records sets.

- **Within-Household Enumeration** -- The data that could be used as a substitute for whole household enumeration also can be used to validate (or supplement) the count and the characteristics data supplied by other modes of enumeration. As the data from other modes are posted to the census data file, they can be matched to the administrative records database to determine whether that information could have been used if one of the traditional modes had not been successful. Again, a set of assignment rules can be applied to help identify households for follow-up.

The administrative records data not accepted for use in a whole household enumeration can be used at the person level to identify people potentially not included at that household during the enumeration process.

- **Content** -- Content items also can be derived from administrative records. Some limited content information can be derived for a subset of addresses. For example, one or more of the test sites may have an assessment record system that carries year built and market value. These data can be used to fill-in the data, to avoid the need for an item nonresponse telephone call or visit, or they may be used in the edit and allocation system, to obtain more accurate results.

## ■ Evaluation Plans

After the census is completed, the Census Bureau staff assigned to the administrative records research project will computer match the administrative records to the 1995 test responses and the responses to the units in the Integrated Coverage Measurement (ICM) sample blocks. Based on the source code in the 1995 test file (identifying in which operation the people were found), staff will determine

whether the administrative records database could have added these people. Nonmatches as a result of this process will be reviewed on two levels: 1) addresses, to determine if they indicate missing housing units and 2) persons, to determine within-household misses.

■ **Access, Privacy, Confidentiality, and Informed Consent**

The administrative records data being sought for the 1995 Census Test are being requested under authority of Title 13, United States Code, Section 6(c), which directs the Secretary of Commerce "... *to the maximum extent possible and consistent with the kind, timeliness, quality and scope of the statistics required, the Secretary shall acquire and use information available from any source referred to in subsection (a) or (b) of this section instead of conducting direct inquiries.*"

Further, all data provided to the Census Bureau from administrative records or the enumeration processes are protected by the Title 13 confidentiality provision (Chapter 7, Section 214), which provides for both a fine and imprisonment for the wrongful disclosure of such information. The protection provided by Title 13, the Bureau's history of protecting confidentiality, and the purely statistical uses of the data (as opposed to any enforcement usage) are key factors in the decision processes used by other governmental agencies deciding to provide their administrative records data to the Bureau. However, the issues of privacy and confidentiality have become broader social issues. These broader issues are likely to have a major negative affect on the Census Bureau's desire to adopt new methodologies, such as this, regardless of the Bureau's record of protecting confidentiality and the privacy of respondents.

The material for this section was developed by Jerry Gates, of the Census Bureau's Program and Policy Development Office; it documents how the issues of privacy and confidentiality reflect on the Bureau.

In 1990, almost four out of five respondents expressed "general concern about threats to personal

privacy in America;" this is up slightly from 77 percent in 1983. However, fewer than two out of three respondents expressed this general concern in 1978[5]. Three surveys by Louis Harris and Associates chart these changes in attitudes: "Dimensions of Privacy" (1978), "The Road After 1984" (1983), and "Consumers in the Information Age" (1990). Professor Alan F. Westin of Columbia University, a leading authority on privacy issues, guided the development and analysis of these surveys.

The survey sponsors deal primarily in insurance and credit. They need information from customers that some customers consider private. These businesses would like to know attitudes towards privacy to help foster customer cooperation. If they fail to respond to customer attitudes, they may go out of business. Therefore, the surveys deal mainly with issues close to the sponsors' interests. However, the surveys ask a few questions important to the Census Bureau.

**Should the government do more to maintain confidentiality of personal information?** In 1978, Louis Harris and Associates asked this question in relationship to eight governmental agencies or types of governments. The respondents gave the best grade to the Census Bureau and the worst to "government welfare agencies."

Governmental unit	Percent		
	Should do more	Doing enough	Not sure
Government welfare agencies	41	35	25
The Census Bureau	26	52	22

In 1978 and 1990, the surveys asked questions related to public confidence or trust in the Census Bureau and other organizations. The questions were different for the two years but the responses follow similar patterns. Most respondents trust the Census Bureau more than other agencies. However, many respondents expressed a degree of mistrust that is



not in proportion with the Census Bureau's record on confidentiality protection. In the table excerpts that follow, the 1990 results allow comparison of the Census Bureau with organizations trusted least and most.

**1978: How confident are you that the Census Bureau protects the privacy of personal information about individuals and does not share it with other government agencies? Number of respondents is 1,511.**

Agency	Very confident	Somewhat confident	Not too confident	Not at all confident	Not sure
Census	14%	32%	26%	19%	10%

**1990: How much do you trust . . .? Number of respondents is 2,254. Respondents rated organizations on a scale from 1 to 10; 1 meant "you don't trust them at all," and 10 meant "complete trust in them." Analysts grouped responses as shown next.**

Type of organization	10-8 High trust	7-4 Moderate	3-1 Low	Not sure
Census	36%	45%	17%	2%
Hospital	37	44	18	1
The IRS	28	39	32	1
Direct mailer	6	28	64	1

It is clear from these data that privacy and concerns about the misuse of data are a concern to many in the public and must be addressed by the Census Bureau if it wants to expand its use of administrative records. It is also true that these types of opinion surveys are sensitive to the form and content of the question being asked. The Bureau has not done research in these areas and needs to do more basic research to better understand these issues.

## ■ Informed Consent

Informed consent is a concept that must be developed further if administrative records are to become a more important source of statistical data. The issues associated with informed consent are several and can be mentioned here only briefly.

- The Privacy Act requires a notice on all Federal forms, but statistical uses of the data are defined as "routine use" that do not require a notice.
- Most answers provided in administrative records datasets are not "optional," so respondents do not have a "choice" about providing such information (other than opting to not participate in the program). This is why the concept of "notice" has evolved. Is notice adequate?
- Some administrative records systems were established before notice was required. How does this affect their use?
- The 1995 Census Test questionnaire will include a "Notice" that informs respondents that their data will be confidential. However, it will not specifically mention the planned use of the answers in administrative records research. The Census Bureau is continuing and intensifying its research in the area of privacy and informed consent. Appendix A is an Executive Summary of a paper, still in progress, dealing with this subject.

## ■ Summary and Next Steps

The 1995 Census Test is the first opportunity to bring together a "complete" set of administrative records for test areas to do a direct comparison with census results. Although events such as the budget approval process or the Census Bureau's inability to acquire key administrative record files may intervene, this test offers the possibility of quantifying the relationship between all existing major -- and

some not so major -- administrative records lists and a census. If successful, the Census Bureau will significantly advance its knowledge about the use of administrative records to reduce the total and differential undercount.

■ **Footnotes**

[1] See Alvey, Wendy and Scheuren, Fritz (1982), "Background for an Administrative Record Census," *Statistics of Income and Related Administrative Record Research: 1982*, Internal Revenue Service.

[2] The sites for the 1995 Census Test are:

Urban sites	Rural site parishes:	
New Haven, CT	Bienville, LA	Natchitoches, LA
Oakland, CA	DeSoto, LA	Red River, LA
Paterson, NJ	Jackson, LA	Winn, LA

Note: Subsequent to the presentation of this paper Paterson, NJ was dropped as a test site due to budgetary constraints.

[3] See Sailer, Peter; Weber, Michael; and Yau, Ellen (1993), "How Well Can IRS Count the Population?" *Statistics of Income: Turning Administrative Systems into Information Systems -- 1993*, Internal Revenue Service.

[4] Census Day for the 1995 Census Test will be March 4, 1995. This "earlier than normal" date was selected to test other factors affecting data collection methodologies.

[5] Equifax, Inc., "Consumers in the Information Age," Equifax, Inc., 1600 Peachtree Street, Atlanta, GA, 30302, 1990, p.1.

## Appendix A

### 'INFORMED CONSENT' AND PRIVACY PROTECTION: Develop- ment of the Doctrine and Applicabil- ity to Census Bureau Activities

*William B. Griffith, in collaboration with R.  
Paul Churchill*

#### EXECUTIVE SUMMARY (6/17/94)

'Informed consent' is an ethical and legal doctrine, the roots of which are to be found in case law (mostly 20th C.), deciding liability claims for unauthorized procedures in medical practice. It emerged as a moral requirement to be imposed on medical research on human subjects, as a response to the shocking revelation in the Nuremberg Trials of the Nazi doctors' experiments. The obligation involves two elements: the subject's agreement to undergo any procedure must be fully voluntary (i.e., free of any coercion, deception, or manipulation); and it must be adequately informed (the standard for this is still developing). It is not well-established as a legal requirement (enforceable by civil actions) in both of these areas, although actual practice in medicine is often criticized as overly formalistic.

The scope of **applicability of this doctrine** of informed consent to social science studies of human subjects, and especially to surveys or studies conducted by interviews, badly needs clarification. The historical evolution of informed consent doctrine, first in medicine and then in social science, particularly social surveying, and its current force as a moral and legal obligation for social data gatherers, forms the **focus of this study**.

The following historical developments are reviewed:

- (1) From its origin in the medical context, informed consent doctrine first crossed over (in

the mid-1960s) into **social and behavioral science research on humans**, as a requirement imposed on most such investigations that received federal sponsorship or funding. Social science researchers in general resisted the imposition of this requirement, minimizing any risks of harms or manipulation of respondents. Federal regulatory authorities then (1981) agreed to exempt most (though not all) federally sponsored or supported survey research from the normal requirement of **prior approval** by an Institutional Review Board. However, this would not eliminate residual tort liability if some provable "harm" resulted, and the question of what counted as a legal "harm" remained open-ended. Although legal enforceability of informed consent on survey research remains somewhat unclear, two national commissions established in the 1970s and 1980s to develop ethical guidelines for all research on humans have emphasized that the requirement for obtaining research subjects' informed consent should be viewed as a significant **moral obligation**, validating an expectation of voluntary conformity, even in the absence of strict legal enforcement.

- (2) Also sketched is a different line of development that contributed to the claim for the doctrine's applicability to social surveying, at least in certain contexts. In the 1960s there erupted a strong surge of public concern over perceived threats to loss of privacy, especially informational privacy, as computerized record systems containing large amounts of personal information proliferated in both the public and private sectors. At the level of principle, the doctrine of informed consent was accepted in the first efforts to formulate a "**code of fair information practices**," in a recommendation that there should be "no use of individually identifiable data...not within the stated purposes... unless the informed consent of the individual has been explicitly obtained" (HEW Report, 1973). In the same period, several Supreme Court decisions re-

interpreted Fourth Amendment protection of privacy, acknowledging it as a "dignity" right of persons, tied to their reasonable expectations, rather than merely to places or papers.

- (3) The study then follows three decades of poorly coordinated and weakly enforced **privacy legislation** to provide individuals some modicum of control over personal information, mostly by requiring notice to respondents of purposes and uses, sometimes by restricting uses. Most of the legislation applied only to federal agencies, depended primarily on agency good faith and/or individual enforcement efforts, and, for the most part, did little to restrain private sector data collectors. Despite clear signals of public concern over developing "surveillance" systems, including polling data and frequent calls for stronger general privacy legislation and enforcement, the law has moved slowly and fitfully to provide legal backing to this moral concern.

The study focuses on two aspects of the present situation:

- (1) There appears to be a significant threat of an imminent eruption over a divergence between (a) a rather widespread public pre-

sumption that individuals have some right to control information files about themselves, and (b) a public perception that personal control is being rapidly eroded with respect to massive, interacting data banks in both the public and private sectors. If such an "explosion" were to occur, an argument that one has observed faithfully informed consent procedures will probably be a salient defense, although all data collectors may suffer harm to some degree.

- (2) The study takes up issues arising from the perceived desirability, in the light of diminishing cooperation by survey respondents, for the Census Bureau to supplement or replace certain information derived from respondents with data derived from existing administrative records. Since consent was not generally sought for such uses at the time of collection, and, generally, it is not feasible to obtain it now, questions arise as to how to protect, in some equivalent way, individuals' interest in having reasonable controls in place when information they have provided is to be reused for a new purpose. The study suggests that appeal to a model more favored in Europe -- e.g. creation of an independent review board -- may be indicated in this context. ■