current population reports

Special Studies
P-23, No. 82

U.S. Department of Commerce

BUREAU OF THE CENSUS

Coverage of the Hispanic Population of the United States in the 1970 Census

A Methodological Analysis

current population reports

Special Studies
P-23, No. 82

Coverage of the Hispanic Population of the United States in the 1970 Census

A Methodological Analysis

by Jacob S. Siegel
and
Jeffrey S. Passel

U.S. Department of Commerce
Juanita M. Kreps, Secretary
Luther H. Hodges, Jr.,
Under Secretary
Courtenay M. Slater,
Chief Economist

BUREAU OF THE CENSUS
Vincent P. Barabba,
Director

Library of Congress Cataloging in Publication Data

Siegel, Jacob S.
Coverage of the Hispanic population of the United States in the 1970 census.

(Current population reports : Special studies : Series P-23 ; no. 82)
1. Hispanic Americans—Statistics.
2. United States—Census, 19th, 1970.
I. Passel, Jeffrey S., joint author. II. United States. Bureau of the Census. III. Title.
IV. Series: United States. Bureau of the Census. Current population reports : Special studies : Series P-23 ; no. 82.
HA203.A218 no. 82 [E184.S75] 312'.0973s
[312'.93] 79-18517

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Postage stamps not acceptable; currency submitted at sender's risk. Remittances from foreign countries must be by international money order or by a draft on a U.S. bank. **Current Population Reports** are sold in two subscription packages: Series P-20, P-23, P-27, and P-60 are available for $40.00 per year ($10 additional for foreign mailing); Series P-25, P-26, and P-28 are available for $70.00 per year ($17.50 additional for foreign mailing). The single-copy price of this report is $2.25.

Preface

As part of its continuing effort to measure and report on the quality of census data, particularly the coverage of major segments of the population, the Census Bureau initiated a study to consider the possibilities of measuring the coverage of the Hispanic population in the 1970 census, at least at the national level. Such a study would complement a previous study which presented measures of the coverage of the White and the Black populations of the United States in the 1970 census. It was recognized early that a study of the kind contemplated would necessarily be exploratory in nature since the lack of comprehensive data on the Hispanic population and the use of various census concepts to identify the Hispanic population would hamper the efforts to arrive at adequate measures of coverage, if not preclude that possibility altogether. In spite of the evident difficulties, the Census Bureau undertook to review the problems and possible methods of evaluating the census count of the Hispanic population and to apply such methods as seemed feasible.

This report represents the results of this effort. It reflects also the Census Bureau's interest in filling the request of the Census Advisory Committee on the Spanish-Origin Population for the 1980 Census and in filling the requirements of Public Law 94-311 (a Joint Resolution of the Congress relating to publication of statistics for the Spanish-origin population), to explore various methods which could be used to produce adequate estimates of the census coverage of the Hispanic population.

This report applies the method of demographic analysis to evaluate the general quality of the data on the Hispanic population in the 1970 census, as well as to measure the coverage of specific age-sex groups. A detailed description of the methods employed in developing the various estimates, including a description of the data and the assumptions incorporated into the methods, is given.

This report supplements two earlier Census Bureau publications presenting estimates of coverage of the population in the 1970 census: "Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis," Evaluation and Research Program of the 1970 Census of Population and Housing, PHC(E)-4, February 1974; and "Developmental Estimates of the Coverage of the Population of States in the 1970 Census: Demographic Analysis," **Current Population Reports**, Series P-23, No. 65, December 1977. The first report presents estimates of the completeness of coverage of the population of the United States as a whole in the 1970 census, for age, sex, and race (White, Black) categories, developed by the method of demographic analysis. The second report presents several alternative series of estimates of the coverage of the population of States in the 1970 census, representing the results of an exploratory effort to apply the demographic method for measuring geographic variations in coverage.

Jacob S. Siegel, Senior Statistician for Demographic Research and Analysis, Population Division, and **Jeffrey S. Passel**, Demographic Statistician on the Research and Analysis Staff of the Population Division, are responsible for the preparation of this report and the research underlying it. Siegel initiated and directed the study, while Passel designed the particular calculations and prepared the basic draft of the text. **Janet Kalwat** served as a professional assistant on the project during a summer internship. **Rita A. Daly** and **Gary D. Smith** assisted the professional staff in carrying out the various calculations. **Mary J. Kisner** typed the various drafts of the report with the assistance of **Joan M. Kans.**

The provisional draft of the report was made available for comment in advance of publication to several social scientists with special knowledge and interest in the areas of census evaluation and Hispanic statistics, namely **Harley Browning, Leobardo Estrada, David C. Heer, Guillermina Jasso,** and **Julian Samora.** The authors wish to thank those who responded to the request for comments.

The Bureau of the Census would also like to thank the Census Advisory Committee on Population Statistics, the Census Advisory Committee on the Spanish-Origin Population for the 1980 Census, the General Government Division of the U.S. General Accounting Office, and the Subcommittee on Census and Population of the Committee on Post Office and Civil Service, U.S. House of Representatives, for encouraging the Bureau to continue its research on the subject treated here and to prepare an appropriate report. The Bureau of the Census is wholly responsible, however, for the contents of this report, including the analyses and interpretations as well as the selection of materials.

This report has raised a number of unresolved problems and has, in effect, been essentially exploratory. Readers' comments and suggestions regarding its contents and possible approaches to the problems raised are invited.

Contents

	Page
Introduction and Summary	
Introduction	1
Summary	2
Methods of Coverage Evaluation	
Case-by-case matching	4
Reinterview survey	4
Record checks	4
Demographic methods	5
Aggregated data from administrative records	5
Demographic analysis	5
Statistical methods	5
Alternative Definitions of the Hispanic Population and Implications for Coverage Evaluation	
Alternative definitions in census data	6
Country of birth or parentage	6
Mexican race	7
Spanish surname	7
Spanish language	7
Spanish heritage	8
Spanish origin	8
Implications of alternative definitions for coverage evaluation	8
Differences in census counts	8
Consistency of response	9
1970 Content Reinterview Survey	10
CPS longitudinal match studies	11
Special census of Gallup, New Mexico	11
Spanish origin-Spanish surname cross-tabulations	12
National Content Test	13
Relation to measurement of coverage	14
Applications of Coverage Evaluation Methods	
Data on the Hispanic population useful in coverage evaluation	15
Census data	15
Non-census data	15
Vital statistics	15
Immigration data	16
Administrative record data	16
General indicators of errors in age and sex data	17
Heaping indexes	17
Age ratios	18
Age-sex accuracy indexes	19
Census survival ratios	20
Native population of Mexican parentage	21
Native population of Puerto Rican parentage	22
Population of Puerto Rican birth or parentage	22
Age-specific death rates and life expectation	23
Sex ratio analysis	25

Contents—Continued

Page

Specific measures of net errors in age-sex data. 29
Comparison with expected populations 29
Intercensal cohort analysis 30
General considerations 30
Population of Puerto Rican birth or parentage 30
Foreign-born Cuban population 35
Native population of Mexican parentage 38
Stable population models 39
Comparison with aggregate administrative record data 39
Conclusion 40

Prospects for Measuring the Coverage of the Hispanic Population
Prospects for development of data 41
Plans for evaluation studies in 1980 43

CHARTS

Figure
1. Observed and expected sex ratios for the White, Black-and-other-races, and Spanish-origin populations, by age: 1970 26
2. Sex ratios for the Spanish-origin population, by type of Spanish origin and age: 1970 27
3. Sex ratios for various Hispanic populations, by age: 1970 and 1960 28
4. Percent net undercount in 1970 for various subgroups of the Hispanic population as a function of the assumed ratio of undercount rates for the subgroups in 1960 to undercount rates for the Black-and-other-races population in 1960 34

DETAILED TABLES

Table
1. Hispanic population according to various identifiers, for the United States and selected areas within the United States: April 1, 1970 9
2. Consistency of responses to the Spanish-origin question as measured by reinterview studies and the Current Population Survey, by type of Spanish origin: 1970 to 1976 10
3. Proportion of persons of Spanish origin with Spanish surnames, and proportion of persons with Spanish surnames of Spanish origin, for the United States and the five Southwestern States: March 1971 12
4. Proportion of persons of Spanish origin with Spanish surnames, by type of Spanish origin, for the United States and the five Southwestern States: March 1971 13
5. Percent of blended population with ages ending in each terminal digit 0 to 9 and Myers' summary index of heaping, for the Spanish-origin, White, and Black populations of the United States: 1970 17
6. Age ratios and age-accuracy indexes for the Spanish-origin, Spanish-surname, White (corrected), and Black (corrected) populations of the United States, by sex: 1970 18
7. Summary measures of errors in age and sex data for selected Hispanic, White, and Black populations of the United States: 1970 and 1960 19
8. Deviation of census survival ratios for selected Hispanic, White, and Black populations from life table survival rates, by sex: 1960 to 1970 21
9. Age-specific and age-adjusted death rates for the Spanish-surname population of Texas, 1970, and other selected populations, by sex 23
10. Expectation of life at birth, age 20, and age 65 for the Spanish-surname population of Texas, 1970, and selected other populations, 1970 and 1969-71 24
11. Preferred estimates of the percents of net undercount of the population of the United States, by sex, race, and broad age group: 1970 and 1960 29
12. Estimates of bicensal relative error for the population of Puerto Rican birth or parentage, using sample survey data for the age-sex distribution of migrants, by sex: 1960-1970 32

Contents—Continued

Page

13. Estimates of bicensal relative error for the population of Puerto Rican birth or parentage, using census data on migration, by sex: 1960-1970 33
14. Immigrants of Cuban birth, by type of admission, for each year, 1959-60 to 1969-70. 35
15. Illustrative estimates of the percents of net census error for the population of Cuban birth, by sex and age and by sex and year of entry into the United States: 1970. 37
16. Estimates of bicensal relative error for the native population of Mexican parentage, by sex: 1960-1970 38

Introduction and Summary

INTRODUCTION

Much interest has been expressed in an estimate of the coverage of the Hispanic population of the United States in the 1970 census, both for its own sake and for its value in measuring the coverage of the population in specific areas.[1] The Bureau of the Census had considered the feasibility of developing an estimate of the completeness of coverage of the Hispanic population in the 1970 census in connection with its studies of the coverage of the White and Black populations in that census. Estimates of coverage were prepared and published for the total, White, and Black populations (with age and sex detail),[2] but an estimate of the coverage of the Hispanic population could not be prepared at the same time because of the lack of appropriate data and an acceptable methodology. This difficulty was complicated by the lack of a consistent and definitive identifier for the Hispanic population. In this report, these obstacles to the estimation of the coverage of the Hispanic population in the 1970 census are explored more fully and possible strategies for overcoming the obstacles are discussed.

Any estimate of "coverage" (net census error) for a particular subgroup of the total population, such as the Hispanic population, may be viewed as consisting of two components: net coverage error and net classification error. With reference to the Hispanic population, net coverage error is the excess of persons of Spanish ancestry omitted from the census over persons of Spanish ancestry counted twice or erroneously included in the census. Net classification error is the balance of persons erroneously classified as of non-Spanish ancestry and persons erroneously classified as of Spanish ancestry. Determining the combined error or the separate components for the Hispanic population is a major problem. In the 1970 census, which relied largely on self-administered questionnaires and self-identification of ethnic affiliation, the measurement of classification error is especially problematic. Such errors may arise from a number of sources. They may result from misinterpretation of the question on ethnic affiliation on the part of the respondent or enumerator, misreporting of ethnic affiliation by the respondent (e.g., born in Cuba, but reporting "not Hispanic"), or processing errors. They may also be associated with inconsistent application of the various definitions of the Hispanic population on the part of the respondent or enumerator or variability in response to the same identifier.

The estimation of net census error rates for any population group requires adequate data for the construction of an "expected" population, with which the census figures can be compared, or reinterview or administrative records covering all or most of the population, with which the census records can be matched. Most of the data available on the Hispanic population come from the decennial censuses or the Current Population Survey (CPS), the Census Bureau's continuing national sample of the population. A small part of the data comes from State and local sources; these data are generally limited to the corresponding geographic areas and to recent years. Administrative record data in which members of the Hispanic population are identified are rare. Where they exist, restricted geographic coverage and population coverage limit their utility for evaluating census data.

Evaluation of the coverage of the Hispanic population in 1970 is complicated in particular by the fact that there may have been a substantial (but unknown) number of illegal residents of Hispanic origin in the United States at that time. Accordingly, some of the "findings" presented here regarding the coverage of the Hispanic population and, in particular, of the population of Mexican origin must be qualified by the fact that no direct allowance was made in the analysis for illegal immigration prior to 1970.

This report includes (1) a discussion of the principal methods that can be used to evaluate population coverage in the 1970 census and the general advantages and limitations of each for the evaluation of the coverage of the Hispanic population, (2) a survey of the census data available on the Hispanic population to be evaluated, (3) a discussion of the implications of the alternative definitions of the Hispanic population in the census for coverage evaluation, (4) a survey of the demographic and administrative record data available

[1] The term "Hispanic" is employed here as a general term for the population of Spanish background or ancestry, and is used variously to encompass any of the several specific concepts which have been used to identify this group on the basis of specific questions or measures and which are quantitatively represented in Bureau of the Census tabulations. The term does not, therefore, have any specific statistical or numerical connotation.

[2] U.S. Bureau of the Census, 1970 Census of Population and Housing: Evaluation and Research Program, PHC(E)-4, **Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis**, 1974.

on the Hispanic population useful for evaluating coverage, (5) a description of the specific procedures of demographic analysis employed here to evaluate the accuracy of the 1970 census data on the Hispanic population and a discussion of the illustrative results, and (6) consideration of possible alternative approaches and possibilities for 1980 and beyond.

SUMMARY

The methods available for estimating the coverage of a population group such as the Hispanic population may be grouped under three general categories: case-by-case matching (employing reinterview surveys or record checks), demographic methods, and statistical methods. The possible application of these methods is explored in this report. The applicability of these methods to the Hispanic population is dependent on the availability of appropriate data from census and other sources. Our present belief is that, of the methods discussed, a national reinterview survey designed for the specific purpose at hand, supplemented by a check against administrative records, could possibly provide the most satisfactory direct information on the coverage of the Hispanic population. Such a survey was not conducted following the 1970 census, however, partly because funding was not available for such an activity at that time. Other survey data and demographic analysis provide the basis for the present exploratory efforts to measure coverage of the Hispanic population in the 1970 census.

Attempts to measure net census error in 1970 for the Hispanic population were severely limited by two problems: the variability of response to the several classifiers or identifiers and the paucity of appropriate data. A central problem is the inability of the census data to reflect a clear, unambiguous, and objective definition of exactly who is a member of the Hispanic population. The Census Bureau has tried several different ways of identifying the Hispanic population in past censuses, but each identifier presents serious problems for use in measuring the national coverage of the Hispanic population. In 1970, four different identifiers of the Hispanic population were used: Spanish origin or descent, Spanish surname (five States only), Spanish mother tongue, and Spanish birth or parentage. These identifiers were then used to define six different populations: Spanish origin or descent, Spanish surname (five States only), Spanish language, Spanish heritage, Spanish language or surname, and Spanish birth or parentage. The primary difficulty with most of these identifiers and classifiers is the amount of variation in response associated with them. It is possible, in fact, that the differences between the counts of the Hispanic population according to the various identifiers and classifiers are larger than the coverage error of the Hispanic population. Furthermore, a number of studies show a substantial amount of variability in individual responses over time. The implications of these definitional problems for the estimation of coverage of the Hispanic population in 1970 are explored in this report.

The evaluation of census data by demographic techniques requires data from sources other than the census. The usual non-census sources—vital registration, immigration records, and administrative records— provide only limited data on the Hispanic population. The data that are available are either incomplete or are of unknown completeness in terms of either population coverage or geographic coverage or do not adequately identify the Hispanic population. As a result, they may not be compatible with any identifier from the census or may require major supplementation (as, for example, for illegal immigration or understatement of emigration). The lack of satisfactory national data on the Hispanic population from independent sources means that the analyses to the coverage of census data which can be made by conventional demographic techniques are mainly checks on the internal consistency of the census data and provide only rough general indications of errors rather than specific estimates of coverage.

Among the techniques providing general indications of error which have been applied to the census data on the Hispanic population for 1970 are the measurement of age heaping, use of indexes of age-sex composition, and the analysis of census survival ratios, death rates, and sex ratios. As a result of applying these techniques no gross errors were found in the data.

Some of the standard demographic techniques for measuring coverage, such as the use of survivors of births as the expected population for a substantial part of the age distribution, were not applicable to the Hispanic population. Applying these techniques would have required making unsupportable assumptions of such broad scope that the resulting estimates of coverage would be conjectural at best. One approach, intercensal cohort analysis, did yield rough estimates of the change in coverage of three selected Hispanic subgroups between the 1960 and 1970 censuses, along with some illustrative estimates of net census error for 1970. Primarily on the basis of these calculations, some general observations relating to the coverage of the Hispanic population in 1970 can be made. The coverage of Hispanic males in 1970 was substantially worse than the coverage of Hispanic females and coverage of young adults was substantially worse than the coverage of other age groups. The population of Cuban birth showed an overall undercoverage of about 3 to 4 percent in 1970. Unlike the White and Black populations, coverage of the Puerto Rican population and the second-generation Mexican-American population appears to have improved substantially between the 1960 and 1970 censuses. The estimates suggest roughly a 5-percent improvement for the population of Puerto Rican birth or parentage and a 6-percent improvement for the native population of Mexican parentage (excluding age cohorts born during the decade, i.e., under 10 years of age in 1970). Coverage of these three subgroups of the Hispanic population, taken as a group, appears to have improved by about 5 percent between 1960 and 1970. However, the three groups, constitute less than two-fifths of the reported population of Spanish origin in 1970.

Although intercensal cohort analysis provided indications of the 1960-70 changes in coverage, the lack of coverage estimates for the 1960 census and of comprehensive vital statistics and immigration data for the Hispanic population for the 1960-70 period precluded the calculation of definitive estimates of the absolute level of coverage of the Hispanic population in the 1970 census. These difficulties limited the development of coverage estimates in 1970 to ones which could be designated as illustrative at best.

The results of demographic analysis suggest a tentative hypothesis regarding the relative coverage levels of the Hispanic, White, and Black populations, that is, that the coverage of the Hispanic population in the 1970 census was intermediate between the coverage of Whites and the coverage of Blacks. The general indications of errors, such as those provided by indexes of age heaping, age ratios, and sex ratios, are consistent with this hypothesis. It is further supported by intercensal cohort analyses which yielded intermediate coverage estimates for three selected subgroups of the Hispanic population (foreign-born Cubans, Puerto Ricans, natives of Mexican parentage) in 1970 for a wide range of assumptions as to the coverage rates for these subgroups in the 1960 census. This type of analysis could not be carried out for the foreign-born population of Mexican origin and, hence, any "finding" regarding the coverage of the Hispanic population based on calculations for the three subgroups fails to incorporate the effect of the possibly differential coverage of several numerous foreign-born subgroups, which may have included a substantial number of persons illegally resident in the United States.

The possibilities and prospects for estimating the coverage of the Hispanic population in 1980 are also considered. Successful application of demographic techniques to this problem will mainly require the development of sources of data on the Hispanic population other than the census. Steps are being taken to expand the range of Hispanic data, e.g., vital statistics, but these will be of little utility for 1980. Resolution of the problem of the subjectivity of the identifier and the inconsistency in the identification of Hispanic persons would also aid greatly in developing adequate coverage estimates, but the prospects here are for only limited improvement. Case-by-case match studies, such as a Census/Post-Enumeration Survey/Social Security Match Study, are being planned for 1980 by the Bureau of the Census. If such studies are successfully carried out, then the prospects for estimating the coverage of the Hispanic population in the 1980 census should be considerably better than they were for the 1970 census.

Methods of Coverage Evaluation

A variety of methods have been devised to evaluate census data. The Census Bureau has used many of these methods to measure the coverage of the population in recent censuses. These methods are enumerated and briefly described here, with some indication of the issues and problems in applying them to the evaluation of the coverage of the Hispanic population in the 1970 census. These methods may be classified as (1) case-by-case matching, employing data from either (a) a reinterview survey or (b) records; (2) demographic methods, employing either (a) demographic analysis or (b) comparison with aggregated data from administrative records; or (3) "statistical" methods, involving either (a) synthetic methods or (b) correlation estimation methods.[3]

CASE-BY-CASE MATCHING

Reinterview Survey

A reinterview survey consists of reenumerating a probability sample of households and matching the individuals in them on a case-by-case basis with the census, for the purpose of checking the coverage of the population represented by the households in the census (or the accuracy of the reporting of the characteristics of the matched persons). Two major limitations of the reinterview method are its requirement of perfect or nearly perfect matching and the tendency for coverage errors in the reinterview to be correlated with coverage errors in the census (i.e., the chance of being excluded from the census is associated with the chance of being excluded from the reinterview). The first limitation tends to result in an overstatement of the omission rate, the second in an understatement. These limitations affect the use of this method for evaluation of the coverage of the Hispanic population as well as the total population. The tendency of reinterview studies to suffer from correlation bias, in particular, sharply limits their utility for the direct estimation of the coverage of the Hispanic population.

A principal advantage of the reinterview procedure is that the coverage of the reinterview survey does not have to be complete to establish the true level of undercoverage in the census.[4] Another advantage is its ability to measure the components of net census error (that is, to distinguish coverage error from reporting or classification error) if the appropriate match studies are carried out. Because of the different concepts employed in the census to count the Hispanic population, the capability of the reinterview procedure to measure the principal components of error is especially useful. The in-depth probes generally included in the reinterview, such as alternative and detailed forms for questions, provide valuable information relating to group definition.

Shifts in identification of individuals as Hispanic over time with the same identifier, differences in identification of individuals as Hispanic with different identifiers for the same date, and differences in the identification of individuals as Hispanic in synchronous surveys, such as a census and a reinterview sample survey, with the same identifier can be measured and analyzed by use of a sample reinterview survey. The Census Bureau has conducted a number of studies using techniques similar to the reinterview method that were designed to measure such shifts and differences. The specific studies are discussed in later sections of this report.

Record Checks

A record check consists of matching a list of names, either a sample drawn from a set of records or the complete set, against the census being evaluated. The principal advantage of the record-check method over the reinterview method is that omissions from the record file are less likely to be correlated with omissions from the census. As with the reinterview method, completeness of the list is not a necessary condition for its use in evaluating coverage in the census; the two collection systems need only be independent, that is, the chance of inclusion in each of the two systems should not be correlated. Moreover, these methods do not require use of other external data such as migration data and, hence, they can be employed effectively to measure coverage for geographic areas within the United States.

In practice the limitations of the record-check method are the same as for the reinterview method. The matching problem remains and, in spite of the greater likelihood of

[3] For a fuller explanation, with illustrative applications, see U.S. Bureau of the Census, **Current Population Reports**, Series P-23, No. 56, "Coverage of Population in the 1970 Census and Some Implications for Public Programs," August 1975, especially pp. 1-13; and Series P-23, No. 65, "Developmental Estimates of the Coverage of the Population of States in the 1970 Census: Demographic Analysis", December 1977, especially pp. 1-9.

[4] See Eli S. Marks, William Seltzer, and Karol J. Krotki, **Population Growth Estimation**, The Population Council, New York, 1974, esp. Chapter 2.

independence between the census and the record file, the two sets of records will not be completely independent. While the record is accessible at any time for matching of individuals, the form and content of the record file must be accepted as given; as a result the matching procedure is rendered more difficult. Furthermore, because the analysis depends on the content of the record files, the record-check method tends to provide less information on the components of net census error.

The most serious problem in the application of the record-check method for evaluating the national or regional coverage of the Hispanic population in the 1970 census is the lack of an adequate set of records, other than the census, which is national or regional in scope and in which the Hispanic population is identified. This situation precludes the direct application of the record-check technique to the measurement of the coverage of the Hispanic population for 1970.

DEMOGRAPHIC METHODS

Aggregated Data From Administrative Records

Estimates of net census errors can sometimes be obtained by comparing census data with aggregated data from administrative records such as Social Security records of covered workers, Medicare records, or school enrollment records. The administrative record file must be complete or must be adjusted for incompleteness, and further adjusted for differences in scope and definition from the census. This method can provide estimates only for particular age-sex segments of the population and for the net census error, not the (coverage and classification) components of error. The major shortcoming of this method for the estimation of the coverage of the Hispanic population is, again, the lack of a set of administrative records identifying the Hispanic population with which the census can be compared.

Demographic Analysis

The method of demographic analysis consists essentially of the development, by various demographic techniques, of expected values for the population in the census categories to be evaluated and the comparison of these expected values with the actual census counts. The expected values are derived by combining data essentially independent of the census being evaluated, such as birth, death, and immigration statistics and data from other censuses, and employing such techniques or devices as life tables, intercensal cohort analysis, expected sex ratios, and population models. Demographic analysis can provide an estimate of net coverage error for the entire population (for which there is no classification error) and estimates of net census error, which combine both coverage and reporting error, for specific groups in the population, such as age, sex, and race groups. The principal limitation of demographic analysis for estimating census coverage is that the expected population developed by this method is directly affected by errors in the basic demographic data and the methodological assumptions employed.

Because of the lack of data on the Hispanic population from independent sources for the entire country or for a particular region or regions, the possible application of demographic analysis to the estimation of the coverage of the Hispanic population in the 1970 census is severely limited. Rough indications of the overall quality of the data for the Hispanic population can be secured, however. The demographic techniques employed do not provide actual estimates of coverage error; rather, they give indications of internal inconsistencies in the census data that can be used to suggest the occurrence of coverage and reporting problems, particularly in certain age-sex groups. One technique, intercensal cohort analysis, is also used to derive some rough estimates of bicensal relative coverage error in the 1970 and 1960 censuses for several subgroups of the Hispanic population.

STATISTICAL METHODS

Another possible approach to the estimation of the coverage of the Hispanic population for subnational areas in the 1970 census involves the application of "statistical" methods. There are many variations of these methods but the primary variations are the synthetic and correlation-estimation methods. The standard version of the synthetic method involves the application of rates or proportions, for specific segments of the population (e.g., socioeconomic or residence categories), relating to some characteristic of the population (e.g., coverage) at a given geographic level (e.g., the United States), to the population at some subordinate level (e.g., States). For example, synthetic coverage estimates for States could be derived by applying national coverage rates for income classes to State populations disaggregated by income. Synthetic estimation could be applied to the Hispanic population if detailed coverage estimates for demographic and socioeconomic segments of the national Hispanic population, or even of the national total population, were available. This requirement cannot be met for 1970. The correlation-estimation method requires detailed estimates of coverage for at least a sample of geographic subdivisions of the United States and hence cannot be applied in 1970.

Alternative Definitions of the Hispanic Population and Implications for Coverage Evaluation

ALTERNATIVE DEFINITIONS IN CENSUS DATA

Recent censuses have provided some data on the Hispanic population. The Bureau of the Census has used a number of different bases to identify the Hispanic population or some of its subgroups in decennial censuses. The various identifiers include:

a. Country of birth and country of birth of parents (1880 to 1970)
   1. Mexico, Cuba, Central or South America, other (1880 to 1970)
   2. Puerto Rico (1950, 1960, 1970)

b. Mexican "race" (1930 only)

c. Spanish surname in five Southwestern States (1950, 1960, 1970)

d. Spanish mother tongue (1940 and 1970)

e. Spanish origin, by type: Mexican, Puerto Rican, Cuban, Central or South American, other (1970)

By combining the four identifiers employed in 1970, the Bureau of the Census defined six different Hispanic populations either regionally or nationally: Spanish origin or descent, Spanish surname (five States only), Spanish language, Spanish heritage, Spanish surname (five States only) or language, and Spanish birth or parentage.

Problems of comparability, coverage, and insufficient scope limit the utility of these data in evaluating the 1970 census.[5] Each of these identifiers presents special problems in the precision with which the Hispanic group is defined, the degree of applicability of the identifier to the Hispanic population, the difficulty of constructing an expected population, and the utility of the identifier for estimating coverage.[6]

[5] U.S. Bureau of the Census, "Data on the Spanish Ancestry Population Available from the 1970 Census of Population and Housing," **Data Access Descriptions**, DAD No. 41, May 1975, and **Persons of Spanish Ancestry**, Supplementary Report PC(S1)-30, 1970 Census of Population, February 1973.

[6] José Hernández, Leo Estrada, and David Alvírez, "Census Data and the Problem of Conceptually Defining the Mexican-American Population," **Social Science Quarterly**, Vol. 53(4), March 1973, pp. 671-687.

Country of Birth or Parentage

The questions on place of birth of each individual and country of birth of parents have provided the only consistent method of identifying foreign stock from Spanish-speaking countries in successive censuses. These questions are currently of limited utility in identifying members of the Hispanic population, however, because there are large numbers of people of Spanish (particularly Mexican) origin who are third-or-higher-generation residents of the United States. In some areas of the Southwest, particularly New Mexico, this problem is especially acute because there are many people of Spanish origin with a distinctive Mexican culture whose ancestors have lived in the same area for centuries.

For some of the other Spanish-origin subpopulations, in particular the Cuban and Puerto Rican populations, most individuals are first-or-second-generation residents, so that data on country of birth and country of birth of parents should cover almost all persons of this background. For these groups very rough coverage estimates might be obtained by demographic analysis using immigration statistics classified by country of origin for the last few decades from the Immigration and Naturalization Service (INS) and estimates of births and deaths to this group for this period. (See the subsequent section, "Intercensal Cohort Analysis.") The reliability of these coverage estimates would still be in question, however, because vital statistics are not available according to country of birth or country of birth of parents. Place of birth of parents is recorded on the birth certificate, and place of birth is recorded on the death certificate, but this information is not tabulated by the National Center for Health Statistics.

The population of Puerto Rican birth or parentage in 1970 as well as in 1960 and 1950 may be identified by the questions on place of birth and place of birth of parents. In these censuses, these classifiers identified practically all of the population of Puerto Rican origin because even in 1970 this population still consisted mainly of first or second generation residents in the United States. (Less than 10 percent of the second generation of Puerto Ricans in the United States was over 30 years of age in 1970.)

Mexican "Race"

The Bureau of the Census tried to identify all persons of Mexican origin, regardless of generation, in the 1930 census by including a category "Mexican" in the question on race. The classification of Mexican as a race was markedly unpopular, particularly with the Mexican Government, and was not attempted again.

Spanish Surname

In 1950, 1960, and 1970, the Census Bureau attempted to identify Mexican-Americans in five Southwestern States (i.e., Arizona, California, Colorado, New Mexico, and Texas) through the use of a list of Spanish surnames. The list used in 1970 was developed by the Bureau of the Census after periodic revisions of a list originally compiled in the 1930's by the Immigration and Naturalization Service. The identification of Mexican-Americans through surnames makes possible the inclusion of persons beyond the first two generations in any tabulation of the group.

This identification technique, as it has been applied, has major shortcomings. Many of the "Spanish" surnames on the list are apparently common among persons of Italian, Portuguese, or similar origins. For this reason, the Census Bureau confined the tabulations based on the list to the five Southwestern States where most of the Mexican-American population lives and where there are few other persons of Latin, non-Hispanic origins (unlike the situation in other parts of the country). Even in this area the correspondence between the population with Spanish surnames and the population of Mexican ethnicity is not always very close. In the five Southwestern States, only about 74 percent of the Spanish-origin population is of Mexican origin and 81 percent of the Spanish-surname population is of Spanish origin.[7] In many parts of these States, however, the problem is less pronounced. In research encoding the city directory of Waco, Texas, according to the Bureau's list of surnames, only 4 percent of the Spanish-surname population could not trace their ancestry to Mexico.[8] On the other hand, as migration into the Southwest by persons from Spanish-speaking countries other than Mexico increases, the likelihood that Spanish surnames will identify Mexican-Americans only will decrease.

A further difficulty with the Spanish-surname classification is that women who marry men with non-Spanish surnames are lost to the Hispanic population, as are their children. About 15 percent of Mexican-American women marry non-Mexican-American men so that the potential loss to the Spanish surname population is substantial. At least part, if not all, of this loss is offset by marriages of Anglo women to Mexican-American men, however.

The geographic restriction to the Southwest places severe limitations on the use of the concept of a Spanish-surname population for estimating coverage in the Southwest by the usual demographic methods. The population is not closed (i.e., is subject to migration) and there is no reliable way of measuring the migration of this population to and from the area. Migrants from the five Southwestern States to the rest of the United States and to foreign countries are lost to the Spanish surname population in later censuses and estimates of the number of such migrants are likely to be very rough at best.

Spanish Language

Spanish mother tongue was used as the principal basis for measuring the Spanish-language population in 1970 but the Spanish-language population included many persons in addition to those of Spanish mother tongue. The Spanish-language population was comprised of persons of Spanish mother tongue and all other persons in families in which the head or wife reported Spanish as his or her mother tongue. The 1970 census question on mother tongue was asked of all persons in the 15-percent sample; mother tongue refers to the language spoken in the person's home when he or she was a child.

Problems of consistency in definition and scope of the identifier complicate the evaluation of the coverage of the Hispanic population based on the Spanish-language identifier. The Spanish-language population is not a closed population and it is not clearly equivalent to the population which considers itself, or is considered by the general population, to be the Hispanic population. Persons may gain Spanish-language affiliation by marriage and lose it by separation or divorce. The children of persons of Spanish mother tongue do not necessarily fall into the Spanish-language population when they became adults. The phenomenon of shifting identification may have been increasing over time as more of the Spanish-language population has assimilated culturally.

It would appear that the Hispanic population cannot be adequately defined by a single characteristic such as language. For example, it is possible that many persons who consider themselves Hispanic may not have spoken the Spanish language at home as a child. As a greater percentage of the Spanish-origin population has become third or higher generation residents and as more and more Hispanic families have moved out of ethnic enclaves, more Hispanic children may have been reared speaking English. As a result, the Spanish-language identifier becomes less successful in identifying the Hispanic population. One final difficulty with this identifier is that it does not differentiate among the various Hispanic subpopulations.

Reliable estimation of the Spanish-language population by demographic analysis is not now possible. In addition to the problem of measuring change due to shifts in identification, there is the problem of measuring the births, deaths, and immigration for this group.

---

[7] U.S. Bureau of the Census, Technical Paper No. 38, **Comparison of Persons of Spanish Surname and Persons of Spanish Origin in the United States, 1975.**

[8] Raymond H. C. Teske, Jr. and Bardin H. Nelson, "Fertility and Related Demographic Variables among Middle-Class Mexican-Americans: A Descriptive Analysis," unpublished paper presented at the 1975 meeting of the Southwestern Social Science Association, San Antonio, Texas, March 1975.

Spanish Heritage

The Spanish-heritage population, defined only for the 1970 census, is an unduplicated combination of populations identified by three of the criteria discussed above, primarily Spanish language. Specifically, it consists of the population identified by:

1. Spanish surname or Spanish language in the five Southwestern States (Arizona, California, Colorado, New Mexico, Texas);
2. Puerto Rican birth or parentage in the Middle Atlantic States (New Jersey, New York, Pennsylvania); and
3. Spanish language in the remaining 42 States and the District of Columbia.

These data have the advantage of being national in scope. However, except for the Puerto Rican population in the Middle Atlantic States, the various Hispanic subpopulations are not distinguished in the figures.

Attempting to evaluate the Spanish-heritage figures by developing an independent estimate of the expected population of Spanish heritage would be an almost impossible task. The population is a loosely concocted mixture rather than a compound of even consistency. Children of members of the Spanish-heritage population may not be members of this population group. Furthermore, individuals may move in and out of the population as they change their residences within the United States. For example, persons of Puerto Rican parentage with a Spanish surname but not of Spanish mother tongue would leave the Spanish-heritage population if they moved from New York to Kansas but would reenter it if they then moved to California. Their children would not be counted as part of the Spanish-heritage population (through Spanish surname) until the families arrived in California. Given problems such as these and the others noted, it would be futile to attempt an evaluation of the 1970 census count of the Spanish-heritage population by demographic methods in spite of the relatively objective nature of the definition.

Spanish Origin

The limitations in the use of Spanish surnames in the Southwest, Spanish country of birth and Spanish country of birth of parents, Spanish language, and Spanish heritage to identify the Hispanic population led the Census Bureau to try another method of identifying this group in the 1970 census, namely, use of a specific question on Spanish origin or descent. In November 1969, a supplement was appended to the Current Population Survey (CPS), which asked for a person's origin or descent (by self-identification) as Hispanic (several categories) or non-Hispanic. Specifically, respondents were asked if they were of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or descent.

This practice has been continued, although the various sets of data are not strictly comparable because the form of the question, the classificatory system and practices, and the sample design have varied. The same question was asked in the 1970 census on the 5-percent sample questionnaire. For the March 1971 and March 1972 Current Population Surveys, the question on origin or descent was expanded to include ethnic groups other than Spanish, and the category "Mexican (Mexicano)." These changes in nomenclature probably had only small effects on the figures for the Mexican-American population as identified. More importantly, the method of assigning children under 14 to the Spanish-origin category was changed in 1973 so that children in households where the wife (but not the head) was of Spanish origin were classified as Spanish. This change added about 300,000 children to the Spanish-origin category between 1972 and 1973.

Two major advantages of the question on Spanish origin as an identifier of the Hispanic population are its ability to identify various subpopulations and its ability to cover third and higher generation residents of Spanish descent. The primary limitation of the Spanish-origin question, particularly for coverage evaluation, stems from the self-designation feature of the question. An individual's conception of himself or herself as of Spanish origin or non-Spanish origin may vary over time.[9] Because of this variability of self-identification, the definition of exactly who is a member of the Spanish-origin population is unclear and often ambiguous. This fuzziness of group boundaries would make the calculation of an independent estimate of an expected population difficult under the best circumstances but the problem is complicated by the lack of adequate data corresponding to this identifier. These issues are particularly relevant for 1980 and later years since a question on Spanish origin or descent is to be included on the 100-percent questionnaire for the 1980 census and self-designated Spanish origin or descent will be the primary means for identifying members of the Hispanic population.

IMPLICATIONS OF ALTERNATIVE DEFINITIONS FOR COVERAGE EVALUATION

Differences in Census Counts

The figures for the Hispanic population of the United States in 1970 vary considerably for the four identifiers (or classifiers) applicable at the national level, from a low of 9,073,000 for the population of Spanish origin or descent to 10,115,000 for the Spanish-language and/or surname population. The second figure exceeds the first by 11.5 percent (table 1). For the five Southwestern States, the area with the greatest concentration of Hispanic population, the relative excess of the highest figure over the lowest figure based on the various identifiers is even greater, about 30 percent. The figures range from 4,668,000 for the Spanish-

[9] Several studies conducted by the Bureau of the Census bear on this point. They are discussed in more detail in the next section of this report.

A similar phenomenon has been found among American Indians; see Jeffrey S. Passel, "Provisional Evaluation of the 1970 Census Count of American Indians," **Demography** 13 (1), August 1976, pp. 397-409.

Table 1. Hispanic Population According to Various Identifiers, for the United States and Selected Areas Within the United States: April 1, 1970

Identifier	United States	Southwestern States[1]	Remainder of United States	Middle Atlantic States[2]	Florida
NUMBER					
Spanish origin[3]	9,072,602	5,008,556	4,064,046	1,749,363	405,036
Spanish surname[4]	(X)	4,667,975	(X)	(X)	(X)
Spanish language[5]	9,589,216	5,662,700	3,926,516	1,873,051	451,382
Spanish heritage[6]	9,294,509	6,188,362	3,106,147	1,052,682	451,382
Spanish language or surname[7]	10,114,878	6,188,362	3,926,516	1,873,051	451,382
Spanish birth or parentage[8]	5,241,892	2,321,642	2,920,250	1,738,802	336,961
PERCENT OF SPANISH ORIGIN[9]					
Spanish origin	100.0	100.0	100.0	100.0	100.0
Spanish surname	(X)	93.2	(X)	(X)	(X)
Spanish language[5]	105.7	113.1	96.6	107.1	111.4
Spanish heritage[6]	102.4	123.6	76.4	60.2	111.4
Spanish language or surname	111.5	123.6	96.6	107.1	111.4
Spanish birth or parentage	57.8	46.4	71.9	99.4	83.2

X Not applicable.

[1]Arizona, California, Colorado, New Mexico, and Texas.

[2]New York, New Jersey, and Pennsylvania.

[3]U.S. Bureau of the Census, Census of Population: 1970, Subject Reports, PC(2)-1C, Persons of Spanish Origin, 1973, table 1.

[4]U.S. Bureau of the Census, Census of Population: 1970, Subject Reports, PC(2)-1D, Persons of Spanish Surname, 1973, table 1.

[5]This group consists of all persons of Spanish mother tongue and all other persons in families in which the head or wife reported Spanish as his or her mother tongue. U.S. Bureau of the Census, Census of Population: 1970, Supplementary Report, PC(S1)-30, Persons of Spanish Ancestry, 1973, table 3.

[6]This group consists of persons of Spanish language or surname in the 5 Southwestern States, persons of Puerto Rican birth or parentage in the 3 Middle Atlantic States, and persons of Spanish language in the remaining 42 States and the District of Columbia. U.S. Bureau of the Census, Census of Population: 1970, General Social and Economic Characteristics, PC(1)-C, 1972, table 85 (United States) and table 49 (States).

[7]Ibid., table 86 (United States) and table 49 (States).

[8]Op.cit., PC(S1)-30, table 5.

[9]Figures represent ratio of specified population to Spanish-origin population (per 100).

surname population to 6,188,000 for the Spanish-heritage population. This difference cannot be attributed to sampling variability since it is considerably greater than its 95-percent confidence interval; it represents, therefore, a real variation in the size of the Hispanic population according to various identifiers.

The large differences between the counts of the Hispanic population according to different identifiers have serious implications for estimating the coverage of the Hispanic population in the 1970 census. If the undercoverage rate of the Hispanic population in 1970 was similar to that of Whites or Blacks, it would fall in the range of 2 to 10 percent. The differences among the various counts of the Hispanic population are generally of this magnitude or larger. Thus, in attempting to measure the undercoverage of the Hispanic population in the 1970 census, we are trying to estimate quantities which are smaller than the differences which can be attributed to definitional variation and, as will be seen next, to response variability for a particular definition.

Consistency of Response

Some evidence is available suggesting that individuals' responses as Hispanic or non-Hispanic vary considerably from time to time. Five studies provide evidence on the reliability and consistency of answers to the Spanish-origin question: (1) 1970 Content Reinterview Survey;[10] (2) CPS match studies for March 1971, 1972, and 1973;[11] (3) special census of Gallup, New Mexico, November 1974;[12] (4) Spanish origin-Spanish surname cross-classifications from the March 1971 CPS;[13] and (5) the National Content Test of 1976.[14]

[10]U.S. Bureau of the Census, 1970 Census of Population and Housing: Evaluation and Research Program PHC(E)-9, Accuracy of Data for Selected Population Characteristics as Measured by Reinterviews, 1974.

[11]U.S. Bureau of the Census, Technical Paper No. 31, Consistency of Reporting of Ethnic Origin in the Current Population Survey, 1974.

[12]Unpublished tabulations and records.

[13]U.S. Bureau of the Census, Technical Paper No. 38, op. cit.

[14]Unpublished tabulations and records.

Table 2. Consistency of Responses to the Spanish-Origin Question as Measured by Reinterview Studies and the Current Population Survey, by Type of Spanish Origin: 1970 to 1976

(Base of percent is the population in the reinterview or in the second-named survey. Nonresponses and responses of "don't know" have been eliminated from the calculations)

Study and measure	All Spanish origins	Mexican origin	Puerto Rican origin	Cuban origin	Central or South American origin	Other Spanish origin	White	Black	American Indian
PERCENT SAME CATEGORY									
1970 Content Reinterview Study[1]....	75.9	(NA)	(NA)	(NA)	(NA)	(NA)	[3]99.5	[3]99.1	(NA)
1971-1972 March CPS Match.....	94.6	91.3	95.0	94.6	71.3	40.9	(NA)	96.2	(NA)
1972-1973 March CPS Match.....	90.9	86.1	92.1	83.3	73.0	46.6	(NA)	96.9	(NA)
1974 Gallup, N.M. Reinterview.....	89.2	74.1	(B)	(B)	(B)	45.9	(NA)	97.9	91.7
1976 NCT Reinterview.....	93.9	96.1	93.9	(B)	(B)	4.6	(NA)	(NA)	(NA)
PERCENT NET DIFFERENCE[2]									
1970 Content Reinterview Study[1]....	+7.9	(NA)	(NA)	(NA)	(NA)	(NA)	[3]+0.3	[3]-0.7	(NA)
1971-72 March CPS Match.....	-3.4	+0.6	+2.9	-10.1	-48.3	-16.7	(NA)	+2.1	(NA)
1972-73 March CPS Match.....	+3.3	+5.7	-2.2	+2.0	-2.5	+5.9	(NA)	+0.2	(NA)
1974 Gallup, N.M. Reinterview.....	+6.0	+15.0	(B)	(B)	(B)	-50.9	(NA)	-2.1	-3.0
1976 NCT Reinterview.....	-4.6	-48.8	-23.5	(B)	(B)	+83.4	(NA)	(NA)	(NA)
PERCENT NON-SPANISH ORIGIN[4]									
1970 Content Reinterview Study[1]....	24.1	(NA)	(NA)	(NA)	(NA)	(NA)	(X)	(X)	(X)
1971-72 March CPS Match.....	5.4	2.7	1.3	2.0	9.2	24.2	(X)	(X)	(X)
1972-73 March CPS Match.....	9.1	4.4	4.4	2.5	18.0	30.3	(X)	(X)	(X)
1974 Gallup, N.M. Reinterview.....	10.8	11.0	(B)	(B)	(B)	9.3	(X)	(X)	(X)
1976 NCT Reinterview.....	6.1	3.6	4.1	(B)	(B)	8.6	(X)	(X)	(X)

B Base of percent less than 40 persons. NA Not available. X Not applicable.

[1]The question in the reinterview differed considerably from the census question. Results may not be comparable with others shown.

[2]A plus sign (+) indicates a net shift into the category in the reinterview phase; a minus sign (-) indicates a net shift out of the category.

[3]From 1970 CPS-Census Match Study.

[4]These figures represent the proportion of a Hispanic subgroup, as identified in the second survey or interview, that was identified as non-Hispanic in the first survey or interview. Most persons reporting different origins in the two interviews reported a Hispanic origin in both interviews.

Sources: Content Reinterview Study--U.S. Bureau of the Census, Census of Population and Housing: 1970, Evaluation and Research Program PHC(E)-10, Accuracy of Data for Selected Population Characteristics as Measured by Reinterviews, 1974, tables A, D, and 1.
CPS Longitudinal Match Studies--U.S. Bureau of the Census, Technical Paper No. 31, Reporting of Ethnic Origin in the Current Population Survey, 1974, tables D and K.
Gallup and NCT Reinterviews--Unpublished tabulations.

Consistency of Response: 1970 Content Reinterview Survey

The Content Reinterview Survey taken following the 1970 census included the question, "Where did your father's (or mother's) ancestors come from?" An individual was then classified as being of Spanish origin or descent if any of his or her ancestors came from a Spanish-speaking country. This classification differs somewhat from the 1970 census classification because it includes as Spanish some people who may think of themselves as primarily of some origin or descent other than Spanish.

The likelihood of reporting Spanish origin in the census declined with the distance of the relationship between the respondents and their immigrant forebears. Of the population born in "Hispanic" countries (according to the reinterview), 99 percent reported being of Spanish origin in the census. For the second generation (at least one parent from a Hispanic country), about 83 percent reported being of Spanish origin in the census; for the third generation, 73 percent; for fourth generation, 44 percent; and for those whose Spanish ancestry is further removed than great-grandparents, only 6 percent reported being of Spanish origin in the census. For persons reporting Spanish origin on both sides of the family in the reinterview, 97 percent reported Spanish origin in the census, whereas only 21 percent with Spanish origin on one side reported themselves to be of Spanish origin in the census.[15]

[15]U.S. Bureau of the Census, Evaluation and Research Program PHC(E)-9, op. cit., pp. 5-8 and table 1.

These data clearly indicate that the relatively objective criterion of having an ancestor, albeit distant, from a Hispanic country does not correspond to the psychosocial criteria applied by respondents to determine their own origin or descent.

Overall, only 76 percent of the persons reporting ancestors from Hispanic countries in the Content Reinterview Survey reported any type of Spanish origin in the census. The consistency of reporting varied considerably according to the characteristics of the respondent. Those with Spanish surnames reported Spanish origin more consistently than those without Spanish surnames. Respondents who were foreign-born reported Spanish origin more consistently than those who were native. Persons living in the Southwest reported Spanish origin more consistently than those in other areas.

Consistency of Response: CPS Longitudinal Match Studies

Because of the design of the Current Population Survey, it is possible to match approximately half of the sample on a case-by-case basis from one year to the next. Such longitudinal match studies were carried out between the March 1971, 1972, and 1973 CPS samples. The consistency of reporting Spanish origin in the March 1971–March 1972 and in the March 1972–March 1973 match studies is summarized in table 2.[16]

These studies indicate that persons of Mexican, Puerto Rican, and Cuban origins report their origin more consistently than persons of Central or South American and Other Spanish origins.[17] Although the populations of Mexican, Puerto Rican, and Cuban origin fall in the range called highly consistent in the report (80 to 95 percent reporting the same category in both years), from the point of view of measuring the coverage of a group even a 5-percent inconsistency is disturbingly and intolerably large. In fact, the 95.0-percent consistency reported for the Puerto Rican origin group in the March 1971–March 1972 match study could be substantially larger or smaller because the 95-percent confidence interval is 90.4 percent to 99.6 percent. The high consistency in reporting of Puerto Rican and Cuban origins is related to the recency of immigration; the vast majority of these persons are first- or second-generation Americans.

For each of the various Spanish-origin populations, a large percentage of those who reported a different origin on the second interview remained within the Spanish-origin population. Many of these individuals chose the "Other Spanish" category on one occasion (giving a local or nationalistic designation such as Hispano or Borinqueño) and a specific origin group on the other. Of those who incorrectly reported the Spanish-origin group, many non-Hispanic persons had misinterpreted the phrase "Central or South American origin" to mean central or southern United States. This problem also affected the 1970 census reports and in addition to distorting the data on Central or South American origin, renders the data for all Hispanic origins combined less satisfactory for estimating the coverage of the group.[18]

For the Nation as a whole, it appears that, from the point of view of estimating the *total size* of the Spanish-origin population, reporting of origin is relatively consistent over time, as is reporting of the major subgroups, Mexican, Puerto Rican, and Cuban. However, from the point of view of measuring the *coverage* of these groups, the amount of inconsistency is excessive. Furthermore, self-identification of Spanish origin is not very consistent over time when specific geographic areas are considered.

Consistency of Response: Special Census of Gallup, New Mexico

In conjunction with the November 1974 special census of Gallup, New Mexico, the Census Bureau conducted an experiment to check the consistency of reporting of Spanish origin. The census questionnaire, which was mailed out, requested the respondent to indicate origin or descent. This mailing was followed up with an interview in which flash-cards were used to solicit a reply to the question on origin or descent. The results comparing the reports on origin or descent are summarized in table 2.[19]

The level of overall consistency for Gallup is similar to the results of the other studies. About 80 percent of the persons calling themselves Spanish in the census did so again in the reinterview. However, the consistency of the Mexican-origin response was quite low—only 74 percent gave the same response. Most of the inconsistent responses involved changes from "Mexican" to "Other Spanish," or vice versa. These shifts illustrate some of the problems in attempting to identify a socially defined group in a self-enumeration census.

A large percentage of the members of the Spanish-origin population in New Mexico often refer to themselves as Hispano or Spanish, and not as Mexican, because they are descended from the original Spanish settlers of the area rather than from immigrants from Mexico. However, the culture of the New Mexican Hispanos is sufficiently similar to that of most Mexican-Americans that the two groups can be combined for most practical purposes.

The Census Bureau has used a number of alternative formats for the Spanish-origin question in order to identify the Hispanic population. Some of the formats included as response categories such terms as "Mexican-American," "Chicano," and "Hispano," either as separate categories or grouped with "Mexican" origin. Including various designations for the Mexican-origin population and tabulating them

[16] Nonresponses and responses of "don't know" in either of the matched surveys have been eliminated from the calculations. Including such responses would, in general, reduce the percentage in the same category, increase the percent net difference, and increase the percent of non-Spanish origin.

[17] Because of sampling variability, other differences between groups may not be significant, however.

[18] The persistence of this problem in tests for the 1980 census has led to the elimination of the "Central or South American" category from the listed responses to the Spanish-origin question.

[19] See footnote 16 regarding treatment of nonresponses.

together increases greatly the consistency of response for this population. Such categories might result in some slight confusion and in the erroneous inclusion of some persons of Portuguese or Brazilian origin as Spanish. However, the costs are likely to be small and are greatly outweighed by the benefits of a more consistently defined population.

The Gallup study highlights another facet of the problem of defining the population of Spanish origin or descent. In certain areas, the terms used by the respondents to define their own group membership differ from the more general terms used by the Census Bureau and other public agencies. This problem is especially prevalent in areas with large concentrations of Hispanic population, such as New Mexico, Texas, California, and New York. To the extent that the distinctive local terms are not recognized by the Census Bureau, the reliability of the data on the Spanish-origin population is lessened and the difficulty of defining a comparable expected population for coverage estimation is increased.

Consistency of Response: Spanish Origin-Spanish Surname Cross-tabulations

Cross-tabulations of the subjectively defined Spanish-origin population and the objectively defined Spanish-surname population provide valuable information on the reliability of both identifiers for delineating the Hispanic population of the United States. Such a study was carried out in connection with the March 1971 Current Population Survey.[20] The Spanish surnames were encoded with the list used for the 1970 census; Spanish origin or descent was tabulated from responses to a question on origin or descent which contained 13 categories, 5 of which were Spanish.[21]

This report concludes that "identification by Spanish surname appears to provide a fair approximation of the Spanish-origin population in the five Southwestern States of the United States, but not in the States outside this area."[22] This relationship can be seen in the data presented in tables 3 and 4. In the five Southwestern States, 81 percent of the population with Spanish surnames identified themselves as being of Spanish origin but, outside these States, only 46 percent did so. Likewise, of the population which identified itself as of Spanish origin, 74 percent had Spanish surnames in the five Southwestern States, but only about 61 percent did in the remainder of the United States. In general, the Spanish-surname and Spanish-origin identifiers correspond better for the Mexican-origin population than for any of the other Spanish subgroups. The same report also concludes that, in spite of the possible losses and gains for the Spanish-surname population through the marriage of women, there is no significant difference between the sexes in the percent with Spanish surnames who report Spanish origin or vice versa.

These findings have important implications for estimating the coverage of the Hispanic population in the 1970 census. Self-declared Spanish origin or descent is the best measure we now have for measuring the Hispanic population. However, nationally, the Spanish-surname population, as measured by the present list of Spanish surnames, does not approximate the Hispanic population as defined by Spanish origin or descent. In fact, about 11 percent of the "Spanish-

[20] U.S. Bureau of the Census, Technical Paper No. 38, Comparison of Persons of Spanish Surname and Persons of Spanish Origin in the United States, 1975.

[21] "Mexicano, Chicano," "Puerto Rican," "Cuban," "Central or So. Amer.," and "Other Spanish."

[22] U.S. Bureau of the Census, Technical Paper No. 38, op. cit., p. 2.

Table 3. Proportion of Persons of Spanish Origin With Spanish Surnames, and Proportion of Persons With Spanish Surnames of Spanish Origin, for the United States and the Five Southwestern States: March 1971

(Numbers in thousands)

Origin and surname	United States	Five Southwestern States[1]	Remainder of United States
Persons of Spanish origin.....	8,957	5,345	3,612
Percent			
Of Spanish surname.....	68.3	73.6	60.5
Not of Spanish surname.....	31.7	26.4	39.5
Persons of Spanish surname.....	9,575	4,850	4,725
Percent			
Of Spanish origin.....	63.9	81.1	46.2
Not of Spanish origin.....	36.1	18.9	53.8

[1] Arizona, California, Colorado, New Mexico, and Texas.

Source: U.S. Bureau of the Census, Technical Paper No. 38, Comparison of Persons of Spanish Surname and Persons of Spanish Origin in the United States, 1975, table A, p. 3.

surname" population outside of the five Southwestern States is actually of Italian origin.[23] Yet the only identifier for which it might be possible to get birth and death data for 1970 and earlier years is Spanish surname. Thus, there is a distinct lack of correspondence between the type of data that can be made available and the population for which coverage is to be measured.[24] Limiting coverage estimates to geographic areas for which Spanish-surname, the identifier available in the census and in non-census sources, corresponds more closely to the Hispanic population requires another type of data which is not available – information on migration within the United States of persons with Spanish surnames.

[23] Ibid., table 9.

[24] Current research at the Census Bureau suggests that it might be possible to develop a list of Spanish surnames which would provide better correspondence with the Spanish-origin identifier and which could be applied throughout the United States. See the section, "Plans for 1980" for further discussion.

Consistency of Response: National Content Test

The National Content Test (NCT) of July 1976 provided another observation on the consistency of reporting of Spanish origin. This survey covered two panels of retired CPS households (each of about 14,000 households). The question on Spanish origin listed seven categories of Spanish origin: "Mexican or Mexicano," "Mexican-American," "Chicano," "Puerto Rican," "Cuban," "Central or South American (Spanish)," "Other Spanish," or "No, none of these." In September 1976, detailed reinterviews were conducted with about 2,300 households from each sample. Respondents in the reinterview phase were asked a series of questions about the ethnicity of parents and ancestors and, finally, a question regarding ethnic self-perception.

The results of this reinterview study, shown in table 2, are generally consistent with the others previously discussed.[25] The reporting of all Spanish origins, Mexican

[25] See footnote 16 regarding treatment of nonresponses.

Table 4. Proportion of Persons of Spanish Origin With Spanish Surnames, by Type of Spanish Origin, for the United States and the Five Southwestern States: March 1971

(Numbers in thousands)

Area and type of Spanish origin	Total persons	Spanish surname	
		Number	Percent
UNITED STATES			
Total, Spanish origin	8,957	6,117	68.3
Mexican	5,023	3,793	75.5
Puerto Rican	1,450	913	63.0
Cuban	626	410	65.5
Central or South American	501	238	47.5
Other Spanish	1,356	764	56.3
FIVE SOUTHWESTERN STATES[1]			
Total, Spanish origin	5,345	3,933	73.6
Mexican	4,358	3,324	76.3
Other Spanish[2]	987	609	61.7
REMAINDER OF UNITED STATES			
Total, Spanish origin	3,612	2,184	60.5
Mexican	665	469	70.5
Puerto Rican	1,407	898	63.8
Cuban	561	384	68.4
Central or South American	406	181	44.6
Other Spanish	573	252	44.0

[1] Arizona, California, Colorado, New Mexico, and Texas.

[2] Includes Puerto Rican, Cuban, Central or South American, and Other Spanish origin.

Source: U.S. Bureau of the Census, Technical Paper No. 38, Comparison of Persons of Spanish Surname and Persons of Spanish Origin in the United States, 1975, table 1, p. 13.

origin, and Puerto Rican origin was reasonably consistent. Over 90 percent of persons reporting one of these origins in the NCT reported the same category in the reinterview. For Mexican and Puerto Rican origins, the net shifts into these categories were extremely large (49 and 23 percent, respectively). Much of this change was the result of shifts of individuals who identified themselves as "Other Spanish" in the original interview into specific origin categories in the reinterview (net shift of 83 percent). The net shifts in the National Content Test appear to be much larger than for some of the other tests. These large shifts can probably be attributed not only to the previously discussed tendency to use local or nationalistic designations, but also to the probing nature of the NCT reinterview.

The National Content Test offers further evidence that self-designation as Spanish is subject to great response variability. The shift between specific categories for individuals designating themselves as Spanish illustrates further the difficulty of developing a set of categories for the Hispanic population which can provide unambiguous identification of individuals as members of specific Hispanic groups.

RELATION TO MEASUREMENT OF COVERAGE

Among other ways, the Census Bureau has been measuring ethnicity on the basis of self-identification of individuals with a group having a common heritage. In 1970, individuals were asked, "What is your origin or descent?" to determine Hispanic ethnicity and were classified as Hispanic if they reported themselves as belonging to a particular Hispanic subgroup. First or second generation residents of the United States can be considered to be of a particular origin or descent whether they designate themselves as such or not. However, individuals of perhaps the third or fourth generation should be counted as of a particular origin or descent only if they consider themselves to be of that origin or descent.

The Hispanic population of the United States, as defined collectively by its members through responses to questions on origin or descent, does not correspond to demographic notions of a population. The offspring of the Hispanic population may or may not be members of the Hispanic population depending on a number of factors, including external circumstances. Membership in the Hispanic population may change from time to time either as a result of change in self-identification or as a result of artificial factors (e.g., responses for other household members including children) which are not closely measurable.

The demographic methods used to estimate coverage are generally based on the assumptions that the offspring of members of a population will also be members of that population and that an individual enters a population by birth or immigration and leaves by death or emigration.[26] These assumptions are obviously not valid for the population defined by the Spanish-origin question in the 1970 census. Thus, conventional demographic techniques are unlikely to give reliable coverage estimates for this population.

Estimation of the net census error of the Hispanic population as defined by origin or descent must take into account both omission of persons and errors in classification. When self-designation is the basis of defining membership in the Hispanic population, the notion of misclassification of adult respondents is, in a strict sense, not applicable since persons are of Spanish origin if they call themselves Spanish and, conversely, are not Spanish if they do not call themselves Spanish. Misclassification can perhaps be deemed to occur for those persons who would classify themselves differently if they responded personally to the question, the question was varied slightly, or the survey was repeated under similar conditions.[27] If sufficient data were available to develop estimates of net census error and its components for the Hispanic population, say from demographic analysis and a post-enumeration survey, it is probable, as suggested by the results of the match studies and consistency tests discussed, that a large part of the estimated net census error would consist of misclassifications of the types noted in addition to coverage error.

The overall impression given by the consistency tests is that the Hispanic population is a socially defined group of persons whose self-identification with the Hispanic population, and particularly with a given Hispanic subgroup, may vary over time and between different records, even without variation in the form of the question. Thus, even if birth and death data could be obtained for the Spanish-origin population, the possible inconsistency in self-designation as Spanish between vital statistics and census data could seriously prejudice the quality of any estimates of coverage error obtained by demographic analysis. The component of the difference between an expected figure and a census figure attributable to changing or inconsistent self-identification could be as large as the component attributable to coverage error. Changes or inconsistencies in self-identification, however, are not properly viewed as census-taking errors; rather such changes or differences reflect the changing views of individuals regarding their relation to other members of society as well as changing sociocultural conditions.

In sum, each of the various identifiers the Census Bureau has used to measure the size of the Hispanic population presents problems in evaluating census coverage for that group. In part, these problems are definitional in character. Lack of a precise definition of the Hispanic population is an evident obstacle to any attempt at deriving coverage estimates for the group by means of demographic analysis. The same problem applies to reinterview and record studies, but in much less degree.

[26] The methods can also be applied to populations where *none* of the offspring are members and net immigration is inapplicable or nonexistent, or where the population is subject to change on the basis of measurable socioeconomic categories such as marriage or divorce.

[27] Misclassification can, of course, also occur as a result of misinterpretation of the question on the part of respondents or enumerators, errors in recording responses, failure to follow the rules for assigning ethnicity to minors, and processing errors.

Applications of Coverage Evaluation Methods

Evaluation of the census count for the total population or an age-sex category by the method of demographic analysis requires an estimate of the true total population or the true population in the age-sex category. Such an estimate is ordinarily obtained by use of data independent of the census under study. In attempting to measure the coverage of the Hispanic population in the 1970 census, a number of demographic techniques were applied with varying degrees of success. The results of these efforts are described in detail following a review of the data useful in evaluating the coverage of the Hispanic population.

DATA ON THE HISPANIC POPULATION USEFUL IN COVERAGE EVALUATION

An adequate evaluation of the coverage of the Hispanic population in the United States in the 1970 census requires a set or sets of data essentially independent of the census itself, national in scope, nearly or wholly complete in coverage, and corresponding in concept to at least one of the identifiers of the Hispanic population used in the 1970 census.

Census Data

Data on the Hispanic population in the 1970 census can be obtained for any one of the six population groups previously enumerated: Spanish country of birth or parentage, Spanish surname, Spanish language, Spanish heritage, Spanish language or surname, and Spanish origin. For comparison with an independently developed estimate of the true population or for estimation of parameters or components used in deriving these estimates, the choice of census data is determined almost solely by the nature of the non-census data on which the estimate is based.

Non-Census Data

Data which can be considered for use in evaluating 1970 census data, particularly by demographic methods, can be obtained from vital records, immigration records, previous census records, Social Security records, and other administrative records.

Non-Census Data: Vital Statistics

Vital statistics tabulations for the Hispanic population are not now available on the basis of any Hispanic identifier from the National Center for Health Statistics (NCHS), the Federal agency which compiles vital statistics. The standard birth and death certificates promulgated by the National Center, however, do contain information which could be used to identify a part of the Hispanic population according to two of the identifiers noted earlier. First, the names (including maiden name of mother on birth certificates) could be coded according to a list of Spanish surnames for the five Southwestern States. The development of vital statistics useful for measuring coverage of the Hispanic population would require coding surnames on birth and death certificates for a substantial period prior to 1970, perhaps 25 to 35 years.

Coding surnames on birth and death certificates for such a period, even if done on a sample basis, is an enormous undertaking and usable data might still not be obtained. Information regarding the completeness of registration of Spanish-surname births or births according to some other Hispanic identifier for the period is required for preparing component estimates (i.e., estimates employing birth, death, and migration data directly). Such information does not exist and cannot be accurately developed. Furthermore, vital statistics for the Spanish-surname population of only five States are of limited utility for demographic analysis in the absence of data on internal migration for the Hispanic population.

Another possibility for developing vital statistics for the Hispanic population, in this case national data, is to code the responses to the place-of-birth question on the death certificate and the place-of-birth-of-parents question on the birth certificate. Such a procedure could be used to identify deaths of first-generation residents and births of second-generation residents of Hispanic origin. Third-and-higher-generation births and second-and-higher-generation decedents could not be identified on this basis. Again, as with Spanish-surname data, these data would be needed for a substantial period of years prior to the 1970 census if they are to be useful for evaluating census coverage in 1970. Such tabulations are not currently available.

Some States and localities which have concentrations of Hispanic population provide vital statistics data for the Hispanic population according to surname, country of origin, or special ethnic identifiers. Three States, Arizona, California, and Texas, code births and deaths according to Spanish surname. Arizona has tabulated births and deaths for the Spanish-surname population since 1969 and Texas

has done so since 1970. California codes its vital records according to Spanish surname but does not tabulate the data. New Mexico has tabulated births and deaths according to ethnic group since 1960, using a classification system which distinguishes Anglo, Spanish, and Mexican. New York State tabulates births and deaths of Puerto Ricans, regardless of place of birth. Some cities and counties also provide tabulations of vital statistics for the Hispanic population: Dallas for Mexican-Americans (based on surnames), New York City for Puerto Ricans, Los Angeles county and San Antonio for the Spanish-surname population, etc. The data for the various States and cities are not all comparable, however. Coding practices and the lists of Spanish surnames employed vary considerably from place to place. The quality of the coding also varies and standards are often lower than acceptable for use with corresponding census data.

The utility of these vital statistics for States and cities for the evaluation of the coverage of the Hispanic population in 1970 is limited. The data are not adequate to develop an expected population even for the specified areas because of the short time span covered and the lack of corresponding data on migration. However, various demographic measures that can provide general indications of the overall quality of the data on the Hispanic population, such as life table measures, can be calculated from the data for some of these areas.

Vital statistics tabulated for Puerto Rico are another useful type of demographic information. Birth and death rates are available for several decades and life tables have been compiled periodically since the early part of the century. These data can serve as bench marks against which various data and measures for the Puerto Rican population and other Hispanic groups in the United States can be compared.

Non-Census Data: Immigration Data

The Immigration and Naturalization Service (INS) provides information on the number of alien immigrants admitted to the United States classified by age, sex, country of birth, and State (or city) of intended residence, as well as a number of administrative categories. These data are available at least for several decades and countries of Spanish language can be identified. This type of data would be useful for deriving an expected national Hispanic population if comparable data on alien emigration, citizen arrivals, citizen departures, and births and deaths were available. These data are lacking in general and cannot be estimated closely enough for use in deriving definitive measures of census coverage for the Hispanic population in 1970. However, illustrative estimates based on the available immigration data are presented in a subsequent section for the population of Cuban birth. Production of these coverage estimates did require a number of unsupported assumptions regarding survival rates, age and sex structure, and emigration because of the lack of suitable data.

Data on arrivals and departures are available for Puerto Rico for each month since January 1959 from the Puerto Rico Planning Board. Prior to 1970 most such movement was destined for or originated in New York. Combined with birth and death statistics for Puerto Ricans in New York City and State, these data constitute the best set of demographic data on any of the Spanish sub-populations independent of the census. Because these data apply only to a local area, however, information on internal migration of the U.S. Puerto Rican population (i.e., net in-migration or out-migration for New York) is also required to develop an estimate of the expected Puerto Rican population for this area in 1970. Such data are not available from sources other than the census, and the census data pertain only to the 1965-70 period as a whole or represent "lifetime" migration. The data on Puerto Rican passenger traffic and census data on 1965-70 migration to and from Puerto Rico have been used to estimate coverage of the population of Puerto Rican birth or parentage for the entire United States in 1970. (See the section, "Intercensal cohort analysis.") The lack of satisfactory survival rates and particularly of satisfactory data on the age and sex structure of the migrants necessitated making a number of unsupported and untestable assumptions which render any coverage estimates for the Puerto Rican population illustrative rather than definitive.

The measurement of the net immigration of the Hispanic population is rendered especially difficult by the possibility of a substantial illegal immigration of persons of Hispanic origin for which there are no reliable data or estimates.[28] The presence in the United States in 1970 of many illegal residents of Hispanic origin who entered in the several years just prior to the 1970 census is widely assumed.[29] However, their number is not known and may be unknowable, except in impractically broad limits. The available material, both analytic and conjectural, supports the view that the number of illegal aliens in the United States in 1970 was substantially smaller than in recent years.

The possibly large numbers of illegal aliens of Hispanic origin in the country in 1970 and the difficulties of estimating the size of this segment of the Hispanic population present serious obstacles to the development of reliable estimates of coverage for the Hispanic population in 1970. This issue will remain for 1980. (See the section, "Prospects for Development of Data" for further discussion of illegal aliens.)

Non-Census Data: Administrative Record Data

There are a few other types of data which could prove useful in estimating the coverage of the Hispanic population. Some States and localities compile data on school enrollment for the Spanish-surname population. Use of these data requires considerable caution because of variations in Spanish-surname lists, coding practices, and coding quality, as mentioned earlier.

---

[28] Charles B. Keely, "Counting the Uncountable: Estimates of Undocumented Aliens in the United States," **Population and Development Review**, Volume 3, No. 4, December 1977, pp. 473-482.

[29] Domestic Council Committee on Illegal Aliens, **Preliminary Report**, December 1976.

Spanish ancestry is not identified in any way in the Social Security, Medicare, or Internal Revenue Service records, but it might be possible to code a sample of the records for the Southwestern States in 1970 according to the Census Bureau's Spanish-surname list. These data could then be used in the aggregate or, more effectively, in a (case-by-case) match study, to evaluate the coverage of the Spanish-surname population in the Southwestern States. The required coding has not been done and would be extremely expensive to carry out at this time. Such a match study could establish whether an individual was included in the census, but it could not establish whether he or she was included in the Spanish-surname count. The limited geographic applicability of this identifier in its present form would still make additional information for other States necessary.

None of the various non-census data which are available meet the criteria set forth earlier for use in evaluation of the coverage of the Hispanic population: independence from the census, national scope, complete or measurable coverage, and agreement with a census identifier of the Hispanic population. Thus, definitive estimates of coverage for the Hispanic population in 1970 are precluded. However, with appropriate demographic analysis almost all of the data can provide some information regarding coverage.

GENERAL INDICATORS OF ERRORS IN AGE AND SEX DATA

The general overall quality of 1970 census data on the Hispanic population can be evaluated through the use of certain limited techniques of demographic analysis even if precise estimates of coverage cannot be prepared. The techniques that have been applied include an examination of the internal consistency of the data, mainly with regard to age and sex reporting. The results of this type of analysis essentially indicate general patterns and types of errors in the data, but they are sometimes useful in identifying specific age-sex groups which are strongly affected by coverage and reporting errors.

Heaping Indexes

Myers' heaping index is used to measure "age heaping," or the tendency of respondents to report ages with certain terminal digits (e.g., zeros, fives) at the expense of others.[30] The method also provides rough minimal indications of the proportion of the population misreporting on each terminal digit. Measures of heaping are considered here because high levels of misreporting in single ages are usually associated with high levels of misreporting of age groups and of omissions. The indexes and proportions are shown in table 5 for the Spanish-origin population in 1970, along with the corresponding figures for the total, White, and Black populations.

Myers' index for the Spanish-origin population in 1970 (0.6) is very low. This value is about the same as the value for the White population in 1970 (0.5) and well below the figure for the Black population (1.5). Myers' index and the proportions misreporting on each terminal digit indicate little or no digital preference in reporting ages on the part of the Spanish-origin population in 1970. In fact, the results

[30] U.S. Bureau of the Census, **The Methods and Materials of Demography**, 2 vols. by Henry S. Shyrock, Jacob S. Siegel, and Associates, third printing, 1975, esp. Vol. 1, p. 207.

Table 5. Percent of Blended Population With Ages Ending in Each Terminal Digit 0 to 9 and Myers' Summary Index of Heaping, for the Spanish-Origin, White, and Black Populations of the United States: 1970

(Percentages were obtained by Myers' blended method, using ages 10 to 79, and should be compared with expected values of 10.0)

Terminal digit	Spanish-origin population			Total population	White population	Black population
	Total	Male	Female			
0	10.1	10.0	10.2	10.0	10.0	10.6
1	9.6	9.5	9.7	9.7	9.7	9.5
2	10.0	9.9	10.0	10.0	10.0	9.8
3	9.9	9.9	9.8	10.0	10.0	9.6
4	10.0	9.9	10.0	9.8	9.8	9.9
5	10.2	10.3	10.1	10.0	10.0	10.2
6	10.0	10.1	9.9	10.0	10.0	9.9
7	10.2	10.2	10.2	10.3	10.3	10.3
8	9.9	10.1	9.8	10.1	10.1	9.8
9	10.1	10.0	10.2	10.1	10.0	10.3
Summary index[1]	0.6	0.7	0.7	0.6	0.5	1.5

[1] Summary index is one-half the sum of the absolute values of deviations from the expected value of 10.0; i.e., summary index = $\frac{1}{2}\Sigma|P_a - 10.0|$.

for the Spanish-origin population are within the range of error of the measurement device and are consistent with no heaping at all.

Age Ratios

More directly relevant to the measurement of coverage is the quality of data for age groups. One method of assessing the reporting and coverage of grouped data on age is to calculate so-called age ratios and an accompanying summary index.[31] Substantial deviations of the age ratios from 100.0 point to age groups which may be overreported or underreported at the expense of adjacent groups and/or which may be underenumerated. The age ratios shown in table 6 for the Hispanic population in 1970, defined in terms of Spanish origin and Spanish surname, identify some age groups that appear to have coverage problems. The corresponding age ratios for the White and Black populations in 1970, corrected for net census undercounts, are shown in the table for comparison.

The 5-9 age group shows ratios greater than 105 for both Hispanic populations reported in table 6. These figures may indicate greater undercoverage of the 0-4 and 10-14 age groups than the 5-9 age group. The very low ratios for males aged 20-24 suggest possibly large net undercounts in this age group. The low ratios for ages 50-54, coupled with the relatively high ratios for ages 45-49 and ages 55-59, suggest substantial age misreporting in this age range, if not also differences in coverage from age group to age group. Likewise, the high ratios for ages 65-69, when considered in combination with the much lower ratios for the adjacent age groups, suggest substantial misreporting into the 65-69 age group at the expense of the adjacent age groups. On the other hand, no firm meaning can be attached to the deviant age ratios because of the distorting effect of past fluctuations in numbers of births and of migration flows into and out of the country. Even the White population corrected

[31] An age ratio is defined here as $\frac{{}_5P_a}{1/3\,({}_5P_{a-5} + {}_5P_a + {}_5P_{a+5})} \times 100$, that is, the ratio of the population in an age group to one-third of the population in the three age groups centered on the numerator population, expressed per 100.

Table 6. Age Ratios and Age-Accuracy Indexes for the Spanish-Origin, Spanish-Surname, White (Corrected), and Black (Corrected) Populations of the United States, by Sex: 1970

(An age ratio is defined as $\frac{3P_a}{P_{a-5}+P_a+P_{a+5}} \times 100$)

Age (years)	Spanish-origin population		Spanish-surname population[1]		White population (Corrected)		Black population (Corrected)	
	Male	Female	Male	Female	Male	Female	Male	Female
5 to 9	105.4	105.2	105.1	105.4	104.0	104.0	103.8	103.6
10 to 14	102.5	101.7	103.6	102.6	103.8	103.5	104.0	103.7
15 to 19	101.4	99.1	104.3	102.2	102.2	100.4	102.6	101.4
20 to 24	94.4	99.8	91.7	98.2	98.6	102.7	94.8	98.5
25 to 29	98.6	97.3	98.2	95.7	100.1	98.4	99.0	97.4
30 to 34	99.2	100.1	97.1	97.6	94.7	94.9	95.8	96.2
35 to 39	99.5	99.5	99.2	99.2	96.7	96.1	99.1	98.6
40 to 44	103.7	102.7	104.0	104.1	101.9	101.5	102.0	101.7
45 to 49	102.6	101.7	105.2	102.5	103.9	104.1	102.4	102.2
50 to 54	92.7	92.7	91.4	93.3	99.9	99.6	97.9	98.0
55 to 59	99.8	101.8	99.0	100.7	101.2	100.6	102.7	102.7
60 to 64	97.4	97.6	99.3	97.6	102.4	102.9	101.2	102.4
65 to 69	105.3	103.5	104.9	106.2	97.3	97.1	95.5	94.1
70 to 74	93.2	94.2	94.8	92.2	98.4	100.9	98.4	101.0
75 to 79	95.8	95.7	91.2	96.2	101.6	103.9	98.2	101.9
Deviations from 100, Irrespective of sign								
Sum of deviations	50.3	40.8	65.2	53.7	35.4	38.4	40.1	37.8
Mean deviation	3.4	2.7	4.3	3.6	2.4	2.6	2.7	2.5

[1] Five Southwestern States only.

for net census errors shows great irregularities in the age ratios and these irregularities may be taken as true reflections of the age patterns of this population.

Age-Sex Accuracy Indexes

Age ratios such as those just described can be combined with one another over the entire age range and with sex ratios in various ways to obtain summary measures of errors in census data on age and sex. These can then be used to assess the general quality of the data. Three such indexes, Das Gupta's age-sex accuracy index,[32] a simple age-accuracy index,[33] and the United Nations' age-sex accuracy index[33] are shown in table 7 for a variety of population groups.

Caution should be observed in interpreting these indexes. They are all based on assumptions of smooth age structures and small, regular changes in sex composition over the age span. The first assumption does not apply very well to the total population of the United States, which has been affected by sharp fluctuations in the number of births over the last several decades. Populations affected by substantial immigration or emigration, such as the population of Cuban origin in 1970 or the population of Puerto Rican birth or parentage in 1960, are especially unsuitable for the application of these age-sex indexes. The native Hispanic population of foreign parentage meets the assumptions of the indexes better than the total Hispanic population (i.e., including the foreign born). At best, the measures identify the levels and variations in errors in census age-sex data only roughly.

The summary indexes of errors in age-sex composition suggest that the 1970 census data for various Hispanic populations are generally somewhat less accurate than the 1970 census data for Whites and Blacks (table 7). Das Gupta's index for the Spanish-origin population in 1970 (2.1) is larger than those for the White population (0.6) or the

[32] Ajit Das Gupta, "Accuracy Index of Census Age Distributions," Proceedings of the World Population Conference, 1954 (Rome), Vol. IV, United Nations, N.Y., 1955, pp. 63-74.

[33] The Methods and Materials of Demography, op. cit., esp. Vol. 1, pp. 222-223.

Table 7. Summary Measures of Errors in Age and Sex Data for Selected Hispanic, White, and Black Populations of the United States: 1970 and 1960

Population	Das Gupta's Grouped Age-Sex Accuracy Index[1]	Age-Accuracy Index	United Nations Age-Sex Accuracy Index[2]
Spanish origin, 1970	2.1	3.0	18.7
Mexican origin, 1970	4.1	4.3	23.3
Spanish surname[3]			
1970	3.0	4.0	21.8
1960	3.4	4.2	21.1
Puerto Rican birth or parentage			
1970	2.7	3.3	22.2
1960	6.2	4.9	29.8
White, 1970			
Enumerated	0.6	2.2	16.0
Corrected	0.7	2.5	17.4
Black, 1970			
Enumerated	0.6	3.2	17.0
Corrected	0.8	2.6	17.6

[1] Das Gupta suggests describing census data as "accurate," "inaccurate," or "highly inaccurate" depending on whether the index is under 5, 5 to 15, or over 15.

[2] The United Nations describes census data as "accurate," "inaccurate," or "highly inaccurate" depending on whether the index is under 20, 20 to 40, or over 40.

[3] Five Southwestern States only.

References: U.S. Bureau of the Census, The Methods and Materials of Demography, by H.S. Shryock, J.S. Siegel, and Associates, Third Printing (rev.), 1975, pp. 218-223.

Ajit Das Gupta, "Accuracy Index of Census Age Distributions," Proceedings of the World Population Conference, 1954 (Rome), United Nations, Vol. IV, 1955, pp. 63-74.

Black population (0.6). However, the value is still quite low and falls well within the range considered by Das Gupta to be "accurate." The index values for the Spanish-surname population and the Puerto Rican population in 1970 are somewhat higher, but are still in the "accurate" range. The Spanish-surname population shows about the same level of accuracy in the 1960 census as in the 1970 census. On the other hand, the population of Puerto Rican birth or parentage shows substantial improvement between the two censuses. In 1960, the population of Puerto Rican birth or parentage was largely of Puerto Rican birth, i.e., it consisted mainly of immigrants. The large proportion of immigrants affected the age-sex structure sufficiently to produce a relatively large value for the summary measure. Under these circumstances, the measure is not wholly indicative of reporting or coverage errors. The factor of immigration probably also accounts for the larger index value for the Mexican-origin population in 1970.

The age-accuracy index shows a general pattern similar to Das Gupta's index. The indexes for the Spanish-origin population and the Black population in 1970 (3.0 and 3.2, respectively) are about the same and relatively small (table 7). This measure suggests less irregularity in the reporting and coverage of Spanish-origin females (2.7) than males (3.4); furthermore, the index for Spanish-origin females is about the same as that for the White female and Black female populations corrected for net census errors. (See table 6.) The United Nations' age-sex accuracy index is slightly higher for the Spanish-origin population (19) than for Whites (16) or Blacks (17) in 1970. However, all three values are below 20 and thus fall in the "accurate" range. The United Nations' index values for the other Hispanic populations reported in table 7 are at or just above 20, with the exception of the much higher value for the population of Puerto Rican birth or parentage in 1960 (30). Again, heavy recent immigration and annual fluctuations in the number of births probably account for the relatively high level of these indexes.

These rough measures of error generally indicate that the data for the various Hispanic populations in 1970 are of fairly good quality. No gross irregularities in reporting or coverage of age groups are apparent in any of the populations. The measures of errors in age data do, however, point to a few age-sex groups in the Hispanic population that may have substantially larger coverage or reporting errors than the rest, in particular, males aged 20-24 and the age groups 50-54 and 65-69.

It is doubtful whether one can draw any clear inferences from these figures regarding the relative accuracy of census data for Hispanics, Whites, or Blacks. In fact, these figures are quite inconsistent with the estimates of relative coverage error for Whites and Blacks derived by direct demographic analysis.

Census Survival Ratios

Analysis of so-called national census survival ratios may be helpful in revealing errors in census age data. Census survival ratios for a population group measure jointly the effects of both mortality during an intercensal period and shifts in net census errors for age cohorts in the two censuses. As such, census survival ratios should be calculated for populations not subject to immigration or emigration during the intercensal period or should be adjusted to exclude net immigration (or to include net emigration) during the period. The computation of census survival ratios for a population requires, at a minimum, data distributed by age for the population from two censuses. For the 1960 and 1970 censuses, the required age data are available for the Spanish-surname population of the five Southwestern States, the population of Puerto Rican birth or parentage, and the first and second generations of Mexican stock.[34]

Adequate census survival ratios cannot be computed even for all of these population groups because of the lack of appropriate data on net immigration (or net internal migration for the Spanish-surname population.) The native population of Mexican parentage can be treated as an approximately closed population (i.e., unaffected by international migration); hence, satisfactory census survival ratios can be computed for it. For the native population of Puerto Rican parentage, adjustments for net immigration can be made; hence, satisfactory census survival ratios can be calculated for this group, too. Appropriate data for making migration adjustments for the Spanish-surname population and for the population of Mexican birth are not available.

The census survival ratios shown in table 8 reflect coverage and age-reporting errors in combination, for age cohorts in the 1960 and 1970 censuses. If both of these censuses were free of coverage or age-reporting errors, the census survival ratios would represent mortality only. They would then exhibit a characteristic pattern similar to that shown by life table survival rates; i.e., all rates would fall below 1.00, the rates would show a smooth decline with increasing age, and the male-female ratios of the rates would be less than 1.00 and shift downward generally with increasing age. (See table 8 for an example based on the 1965 U.S. life tables for Whites and Black-and-other-races.) Deviations from the "expected" pattern, represented by an unweighted average of the survival rates from the White and Black-and-other-races life tables, may be taken as suggestive of problems in the data.

Table 8 shows the deviations of the census survival ratios for the native populations of Mexican and Puerto Rican parentage from the average life table survival rates. The national census survival ratios for the White and Black populations provide a further basis for comparison and analysis of the census survival ratios for the Hispanic populations. Because of the assumptions relating to migration made in the computation of the census survival ratios, differences between the ratios for the Hispanic populations and those for the U.S. life table populations must be interpreted with caution, however.

[34] Tabulations by age and sex are not available from the 1960 census for the first and second generations of other Hispanic populations.

Table 8. Deviation of Census Survival Ratios for Selected Hispanic, White, and Black Populations from Life Table Survival Rates, by Sex: 1960 to 1970

(Differences computed from unrounded figures)

Sex and age (years)		Census survival ratios[1]					Life table survival rates, 1965[3]	Deviation from life table survival rate				
		Native of Mexican parentage	Puerto Rican birth or parentage	Native of Puerto Rican parentage[2]	White	Black		Native of Mexican parentage[4]	Puerto Rican birth or parentage[4]	Native of Puerto Rican parentage[4]	White[5]	Black[6]
In 1960	In 1970											
MALE												
0 to 4	10 to 14	1.225	1.004	1.029	1.002	1.021	.992	+.232	+.012	+.037	+.008	+.031
5 to 9	15 to 19	1.175	.993	1.016	1.005	.999	.992	+.183	+.001	+.024	+.011	+.007
10 to 14	20 to 24	.970	.835	1.083	.987	.912	.985	-.016	-.150	+.098	-.001	-.071
15 to 19	25 to 29	.931	1.271	1.192	.973	.894	.978	-.047	+.293	+.214	-.011	-.078
20 to 24	30 to 34	1.084	1.173	.982	.984	.958	.973	+.111	+.200	+.009	+.000	-.004
25 to 29	35 to 39	1.024	1.054	1.113	.980	.955	.966	+.058	+.088	+.147	-.002	+.004
30 to 34	40 to 44	.974	.967	.980	.972	.939	.953	+.020	+.014	+.027	-.002	+.006
35 to 39	40 to 44	.976	.904	.959	.948	.892	.933	+.043	-.029	+.026	-.010	-.017
40 to 44	50 to 54	.897	.954	.797	.933	.884	.902	-.005	+.052	-.105	+.002	+.010
45 to 49	55 to 59	.884	.901	1.034	.885	.829	.857	+.027	+.044	+.177	-.004	+.004
50 to 54	60 to 64	.857	.925	.934	.845	.819	.797	+.059	+.128	+.137	+.014	+.056
55 to 59	65 to 69	.841	.826	.928	.750	.777	.714	+.128	+.113	+.214	-.004	+.105
60 to 64	70 to 74	.820	.752	(B)	.675	.699	.613	+.206	+.139	(B)	+.016	+.132
65 and over	75 and over	.503	.615	(B)	.389	.379	.390	+.113	+.225	(B)	+.016	-.028
FEMALE												
0 to 4	10 to 14	1.239	.969	1.012	.997	1.015	.994	+.246	-.025	+.018	+.001	+.023
5 to 9	15 to 19	1.179	.942	.998	1.007	1.004	.996	+.183	-.054	+.002	+.010	+.009
10 to 14	20 to 24	.993	.846	1.062	.998	.970	.994	-.001	-.148	+.069	+.003	-.022
15 to 19	25 to 29	.942	1.280	1.340	.989	.994	.990	-.049	+.290	+.349	-.005	+.007
20 to 24	30 to 34	1.095	1.272	1.254	.996	1.028	.986	+.109	+.286	+.268	+.004	+.048
25 to 29	35 to 39	1.060	1.110	1.072	.994	1.009	.980	+.081	+.130	+.092	+.004	+.039
30 to 34	40 to 44	.970	1.004	.959	.989	.970	.970	+.000	+.034	-.011	+.004	+.014
35 to 39	45 to 49	.960	1.019	.954	.968	.914	.957	+.003	+.062	-.003	-.008	-.024
40 to 44	50 to 54	.949	1.038	1.029	.964	.910	.939	+.010	+.099	+.090	+.002	-.006
45 to 49	55 to 59	.949	.982	1.179	.939	.871	.912	+.037	+.070	+.266	-.005	-.010
50 to 54	60 to 64	.897	1.110	1.068	.936	.902	.873	+.024	+.237	+.194	+.017	+.074
55 to 59	65 to 69	.842	1.001	1.108	.891	.928	.815	+.027	+.186	+.294	+.015	+.174
60 to 64	70 to 74	.968	.839	(B)	.837	.808	.743	+.225	+.096	(B)	+.029	+.129
65 and over	75 and over	.684	.684	(B)	.509	.466	.461	+.223	+.223	(B)	+.059	-.006

B Base less than 500.

[1]Male population adjusted to include Armed Forces overseas in 1960 and 1970.

[2]For calculation of census survival ratios, the native population of Puerto Rican parentage in 1960 and 1970 includes persons enumerated in Puerto Rico and born in the United States.

[3]Average of White and Black-and-other-races survival rates.

[4]Standard is average of White and Black-and-other-races survival rates for 1965 (column 6).

[5]Standard is 1965 U.S. White life table survival rates (not shown).

[6]Standard is 1965 U.S. Black-and-other-races life table survival rates (not shown).

Source of life tables: National Center for Health Statistics, Vital Statistics of the United States, 1965, Vol. II - Mortality, Part A, 1967, p. 5-4.

Census Survival Ratios: Native Population of Mexican Parentage

The census survival ratios for the native population of Mexican parentage show patterns roughly similar to those of Whites and Blacks but the ratios for the Hispanic group are more extreme and irregular. The census survival ratios for both male and female natives of Mexican parentage for ages 10-14 and 15-19 in 1970 are much larger than 1.00. These figures can be explained by better coverage of the population at ages 10-19 in 1970 than at ages 0-9 in 1960. This pattern may also be accounted for by an increased tendency for persons born in Mexico to misreport their place of birth as the United States.

The census survival ratios for native males and females of Mexican parentage are well below 1.00 at ages 25-29 in 1970, much greater than 1.00 at ages 30-34 in 1970, and relatively close to 1.00 at ages 20-24. This pattern of census survival ratios suggests lower coverage rates for ages 25-29 than for the surrounding ages in the 1970 census. The particularly high census survival ratios for ages 30 to 39 in 1970 indicate that coverage errors for ages 20 to 29 in 1960 may have been especially serious. Furthermore, the substantially lower census survival ratios for Black males ages 20 to 29 in 1970 compared to the census survival ratios for the Mexican-parentage group could indicate that census coverage was worse for Black males in these ages in 1970 than for Mexican-American males.

Other notable differences occur at the older ages. The census survival ratio for native females of Mexican parentage aged 65-69 in 1970 is too low and the ratio for ages 70-74 in 1970 is much too high, in comparison with the White female life table survival rates for 1965 and the ratios for both White females and Black females at these ages in 1970. This pattern is indicative of some age misreporting—either a preference for ages 70-74 at the expense of ages 65-69 in 1970, or a preference for ages 55-59 at the expense of

ages 60-64 in 1960.[35] Age misreporting is also possibly indicated for native males of Mexican parentage aged 45-54 in 1970; the census survival ratio for males aged 45-49 appears slightly high and the ratio for males ages 50-54 appears slightly low. The other census survival ratios are similar to those for the White population.

The census survival ratios for the native population of Mexican parentage virtually all exceed those of the standard (table 8). The exceptions occur in age groups adjacent to others which exceed the standard by an exceptionally large amount, suggesting age misreporting for these groups. One possible explanation for the general excess in the ratios is that mortality for the Mexican parentage population was lower during the 1960's than in the standard population. Another, more likely, possibility is that coverage improved over the decade more for persons of Mexican parentage than for the total population (approximately 0.3 percent for persons over 10 in 1970). Other evidence supporting this contention is presented in the subsequent section, "Intercensal cohort analysis."

Census Survival Ratios: Native Population of Puerto Rican Parentage

The derivation of the census survival ratios displayed in table 8 for the native population of Puerto Rican parentage required a large adjustment for migration. This adjustment was made, in effect, by combining the population residing in Puerto Rico and reporting the United States as place of birth with the native population of Puerto Rican parentage residing in the United States in both 1960 and 1970. In each year the overseas group was about 18 percent as large as the resident group. Because of the nature of the migration adjustment, the survival ratios for the native population of Puerto Rican parentage are even more problematic than the census survival ratios for the Mexican-parentage population. Deficiencies in the adjustment procedure could easily account for much of the deviations from the expected patterns. Errors in reporting place of birth in the census could also be responsible for some of the deviations. Hence, any inferences regarding the net census errors for the population of Puerto Rican parentage based on the census survival ratios in table 8 are to be interpreted merely as reasonable possibilities.

The high census survival ratios for males of Puerto Rican parentage aged 20-29 and 35-39 in 1970 could indicate coverage errors in the 1960 census at ages 10-19 and 25-29. Similarly, the large ratios for males aged 60-69 in 1970 could indicate relatively large coverage errors in 1960 or considerable age misreporting in both censuses. Age misreporting would seem to be more definitely responsible for the low census survival ratio at ages 50-54 in 1970 in combination with the excessive value at ages 55-59. For females of Puerto Rican parentage the ratios are too large relative to those for Whites and Blacks at ages 20-39 (especially ages 25-34) and at ages 50-69. The remaining ratios for both sexes are similar to the expected patterns.

Taken at face value, the 1960-70 census survival rates for the population of Puerto Rican parentage give an indication of some serious coverage problems for this population in the 1970 or 1960 censuses. As was the case for the native population of Mexican parentage, virtually all the ratios for the population of Puerto Rican parentage exceed those for the standard population. As before, this fact could indicate coverage improvements from 1960 to 1970 in excess of those found in the standard population (0.3 percent).[36] Again, it should be stressed that, because the adjustment for migration is only approximate, these and other indications, by themselves, cannot be treated as definite signs of serious errors in the data or of the absence of serious errors.

Census Survival Ratios: Population of Puerto Rican Birth or Parentage

If we consider the entire Puerto Rican population of the United States, both persons of Puerto Rican birth and persons of Puerto Rican parentage, another possibility for adjusting census survival ratios for migration is to exclude survivors of net migration during the intercensal period from the figures for the second census. The monthly figures on net arrivals and departures between the United States and Puerto Rico can be used to adjust the census survival ratios for the migration component. These data themselves required a substantial amount of adjustment for errors and allocations for unknown characteristics. Unfortunately, the magnitudes of some of the adjustments are sufficiently great that the general patterns and levels of the ratios shown in table 8 could be almost solely attributable to the adjustments themselves rather than to mortality or changes in census coverage.[37] Nonetheless, we shall examine these census survival ratios for what indications of coverage errors they can provide.

The census survival ratios for both sexes at ages 25-39 in 1970 are much too large relative to those for the standard population and the other populations shown in table 8. These large values are partially complemented by ratios which are too small in the preceding age group 20-24. This pattern is suggestive of either age misreporting in one or both censuses or, more likely, errors in the procedures for allocating net migration to age and sex groups. Census survival ratios for Puerto Rican females aged 10-19 in 1970

[35] Age misreporting of both types is likely to have occurred. Heaping indexes computed for other populations show a preference for ages ending in the digit nine in 1960 because the question on age asked only for year of birth. An unusual concentration at age 59, corresponding to year of birth 1900 was further noted. For 1970, the more usual form of age heaping on zeros and fives was again found as the question format included both age and year of birth. Age heaping in recent U.S. censuses has been more pronounced at the older ages so that the pattern of age misreporting hypothesized is consistent with the pattern of survival ratios noted. (See U.S. Bureau of the Census, Census of Population: 1970, Volume 1, Characteristics of the Population, Part 1, United States Summary—Section 2, 1973, p. App-13.)

[36] See the section, "Intercensal cohort analysis", for further evidence supporting this assertion.

[37] The adjustments and the rationale for making the particular adjustments are discussed in detail in the subsequent section "Intercensal cohort analysis: population of Puerto Rican birth or parentage."

are somewhat lower than those for any population group shown in table 8. Better coverage of ages 0-9 in the 1960 census than ages 10-19 in 1970 is not a very appealing or likely explanation as young children tend to be among the age groups with the worst coverage and teenagers among those with the best. However, overstatement of the mortality of these cohorts between 1960 and 1970 or errors in the procedures for allocating migration offer a better explanation for the observed pattern.

At ages over 50 for males and over 30 for females in 1970, the census survival ratios for the Puerto Rican birth or parentage population all exceed those for the standard population. Furthermore, at ages over 60 for both sexes the amount of the excess is large. As was mentioned in the preceding sections, this pattern of excess census survival rates is indicative of coverage improvements from 1960 to 1970 greater than 0.3 percent,[38] if it is assumed that the adjustments for intercensal net migration are accurate or at least approximately correct.

Age-Specific Death Rates and Life Expectation

Analysis of age-specific death rates and life table functions computed from these deaths rates for Hispanic populations can provide some evidence of coverage and reporting errors in the census data for specific age groups as well as indications of the overall quality of the census data. Mortality data for Hispanic populations are not widely available. The Texas State Department of Health provides tabulations of deaths classified by age for the Spanish-surname (and White non-Spanish-surname) population which can be used with the corresponding census data to compute age-specific death rates for the Spanish-surname (and White non-Spanish-surname) population of Texas. Age-specific death rates were computed with deaths of Spanish-surname persons (and White non-Spanish-surname persons) for 1970 as numerators and data for the Spanish-surname (and White non-Spanish-surname) population of Texas taken from the 1970 census as denominators.[39] These rates are shown in table 9, along with age-specific death rates for the White population of the United States in 1970 and death rates from one of the Coale-Demeny model life tables, specifically West model life table, level 22.[40]

[38] Further discussion, including more detailed evidence can be found in the subsequent section entitled "Intercensal cohort analysis: population of Puerto Rican birth or parentage."

[39] The tabulations of deaths were available only in 10-year age groups above age 15; accordingly, the deaths were subdivided into 5-year groups by means of Newton's method (**The Methods and Materials of Demography**, op. cit., pp. 222-223). The population data were taken from U.S. Bureau of the Census, 1970 Census of Population, Subject Reports PC(2)-1D, **Persons of Spanish Surname**, 1973.

[40] Ansley J. Coale, and Paul Demeny, **Regional Model Life Tables and Stable Populations**, Princeton University Press, 1966.

Table 9. Age-Specific and Age-Adjusted Death Rates for the Spanish-Surname Population of Texas, 1970, and Other Selected Populations, by Sex

(Rates per 1,000 population)

Age (years)	Male					Female				
	Texas, 1970		United States, 1970			Texas, 1970		United States, 1970		
	Spanish surname (e_0=67.2)	White, non-Spanish surname (e_0=68.1)	White (e_0=68.1)	Black and other races (e_0=61.3)	West model life table (Level 22) (e_0=68.6)	Spanish surname (e_0=73.4)	White, non-Spanish surname (e_0=76.5)	White (e_0=75.7)	Black and other races (e_0=69.4)	West model life table (Level 22) (e_0=72.5)
Less than 1	31.3	22.3	21.1	40.2	31.6	25.8	15.9	16.2	31.7	23.1
1-4	1.4	1.0	0.8	1.4	1.5	1.1	0.7	0.7	1.2	1.2
5-9	0.6	0.5	0.5	0.6	0.7	0.4	0.4	0.3	0.5	0.4
10-14	0.5	0.6	0.5	0.7	0.6	0.4	0.3	0.3	0.4	0.4
15-19	2.1	1.5	1.5	2.2	1.1	0.6	0.6	0.6	0.8	0.6
20-24	3.4	1.9	2.0	4.2	1.5	0.9	0.8	0.7	1.4	0.8
25-29	3.1	1.4	1.7	4.6	1.5	0.9	0.7	0.7	1.9	1.0
30-34	3.4	2.0	1.9	5.6	1.7	1.3	1.1	1.0	2.5	1.3
35-39	3.5	2.7	2.6	7.2	2.2	2.1	1.5	1.5	4.0	1.7
40-44	4.1	4.4	4.2	10.2	3.2	2.7	2.2	2.3	5.9	2.5
45-49	5.6	7.3	6.8	14.0	5.2	3.8	3.5	3.7	8.3	3.8
50-54	10.5	11.6	11.0	19.4	8.3	6.8	5.3	5.6	11.5	5.8
55-59	16.0	17.8	17.7	26.4	13.6	9.5	7.2	8.3	16.1	8.9
60-64	22.8	25.1	27.1	35.3	21.6	14.5	10.8	12.2	22.2	14.3
65-69	34.3	38.9	40.5	47.6	34.3	24.2	17.1	19.2	31.3	24.4
70-74	55.1	56.6	58.3	65.6	55.0	42.0	27.2	31.3	44.9	42.3
75-79	84.0	82.7	86.9	84.8	88.0	66.5	49.4	53.5	57.8	72.7
80-84	109.6	125.9	126.1	98.6	180.6	92.4	79.2	88.7	74.2	162.0
85 and over	157.9	179.1	185.5	114.1		126.2	160.1	159.8	102.9	
Age-adjusted death rate[1]	11.7	12.0	12.3	15.2	11.9	8.3	6.7	7.2	9.9	9.4

[1] Standard population is 1970 United States total White population.

Sources: Texas, Spanish surname population: Computed from death statistics supplied by the Texas State Department of Health and population data from U.S. Bureau of the Census, Census of Population: 1970, Subject Reports PC(2)-1D, Persons of Spanish Surname, 1973, and General Population Characteristics, PC(1)-B45, Texas, 1971, table 20.

United States: U.S. Public Health Service, National Center for Health Statistics, Vital Statistics of the United States, Vol. II - Mortality, Part A, 1970, 1974, tables 1-8.

Model Life Table: Ansley J. Coale and Paul Demeny, Regional Model Life Tables and Stable Populations, Princeton University Press, 1966.

With few exceptions the age-specific death rates for the male and female Spanish-surname populations of Texas in 1970 are similar (in level and overall pattern) to 1970 death rates for the White, non-Spanish-surname population of Texas and the White population of the United States, and to rates from the West model life table, level 22. Not all of the differences should be taken as indicative of coverage errors. The higher infant death rate for the Spanish-surname population of Texas than for the other populations presumably reflects real differences between the infant mortality of the Spanish-surname population of Texas and the other populations. The relatively low level of the death rates for Spanish-surname males aged 80-84 and Spanish-surname persons of both sexes at ages 85 and over supports the hypothesis that many elderly Mexican-Americans (particularly those born in Mexico) with chronic diseases of later life return to Mexico to spend their last years "at home." On the other hand, the relatively low level of the rates may be an artifact since they are subject to large random fluctuations as a result of the small size of the population involved.

The abnormally high death rates for Spanish-surname males in the age range 15 to 39 may result from the greater incidence of violent and accidental deaths in this population. Another plausible explanation is that these ages suffered large undercounts in the 1970 census which made the denominator of the death rate smaller than the true value and the rate larger. In contrast, the death rates at these and most other ages for the female Spanish-surname population of Texas differ only slightly from those of the other three populations and do not suggest the presence of serious coverage errors among females.

The existence of problems with population data can also possibly be inferred from the corresponding life tables, which summarize the mortality experience implicit in the age-specific death rates for a population. Life tables for the Spanish-surname and White non-Spanish-surname populations of Texas were constructed on the basis of the age-specific death rates shown in table 9 by means of Greville's short-cut method.[4] Measures of expectation of life at birth, age 20, and age 65, taken from these and selected other life tables, are shown in table 10.

The life tables for the Spanish-surname population of Texas in 1970 imply that the mortality experience of this population is roughly equivalent to that of the total White population and the White non-Spanish-surname population of Texas in 1970 and is much better than that of the Black-and-other-races population of Texas in 1970. For example, the expectation of life at birth in 1970 is 70.2 years for the Spanish-surname population, 71.7 years for the (total) White population, and 65.5 years for the Black-and-other-races population. At age 20 similar relationships appear, but at age 65 the groups differ less.

The differences in life expectation between Spanish-surname males and White non-Spanish-surname males of Texas are all quite small: -0.9 year at birth, -0.1 year at age 20, and +0.6 year at age 65 (table 10). Indeed, if some irregularities in the age-specific death rates for Spanish-surname males at ages 15 to 34 are removed, the deficit at birth disappears and the Spanish-surname males have an expectation of life at age 20 which is 0.6 year *higher* than

[4] The Methods and Materials of Demography, op. cit., pp.444-445.

Table 10. Expectation of Life at Birth, Age 20, and Age 65 for the Spanish-Surname Population of Texas, 1970, and Selected Other Populations, 1970 and 1969-71

Population and year	At birth			At age 20			At age 65		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
TEXAS, 1969-71									
Spanish surname[1]	70.2	67.2	73.4	53.2	50.4	56.0	15.3	14.1	16.4
Spanish surname (adjusted)[1 2]	70.5	68.1	73.4	53.3	51.1	56.0	15.3	14.1	16.4
White, non-Spanish surname[1]	72.2	68.1	76.5	54.3	50.5	58.3	15.8	13.5	17.9
White	71.7	67.8	75.9	54.0	50.4	57.8	15.5	13.3	17.5
Black-and-other-races	65.5	61.7	69.5	48.7	45.2	52.3	14.9	13.3	16.5
UNITED STATES, 1969-1971									
All races	70.7	67.0	74.6	53.0	49.5	56.6	15.0	13.0	16.8
White	71.6	67.9	75.5	53.7	50.2	57.2	15.1	13.0	16.9
Black-and-other-races	65.0	61.0	69.0	48.1	44.4	51.8	14.5	12.9	16.0

[1] 1970 only.

[2] Age-specific death rates for ages 15-34 (males) and 35-39 (females) were smoothed graphically.

Sources: United States: U.S. Public Health Service, National Center for Health Statistics, United States Life Tables: 1969-71, 1975.

Texas, White and Black-and-Other-Races: U.S. Public Health Service, National Center for Health Statistics, State Life Tables: 1969-71, 1977.

Texas, Spanish-surname population: Computed from death statistics supplied by the Texas State Department of Health and population data from U.S. Bureau of the Census, Census of Population: 1970, Subject Reports PC(2)-1D, Persons of Spanish Surname, 1973.

for White non-Spanish-surname males. On the other hand, the deficits for females are not negligible at any age: 3.1 years at birth, 2.3 years at age 20, and 1.5 years at age 65. Such differences are probably not caused primarily by coverage errors; they probably reflect real deficits in life expectancy for the most part.

Life tables are available for various "Hispanic" countries and regions, such as Mexico and Puerto Rico. However, the mortality experience depicted in these tables is quite different from that of the Spanish-surname population of Texas. Life expectation at birth in 1970 for Mexico was 61.4 years,[42] well below that computed for the Spanish-surname population of Texas for the same year. For Puerto Rico, life expectation at birth in the 1969-71 life table was 72.0 years, but the distribution of deaths by cause of death in Puerto Rico is very different from the distribution of deaths for the U.S. population of Puerto Rican birth.[43] Thus, the mortality level of the Hispanic population in the United States, as summarized in the life tables for Texas, appears to be more similar to the mortality level of the rest of the U.S. population than to the mortality level in various Hispanic countries of origin.

If we can assume that the deaths and the population used to compute the age-specific death rates represent a common population and *if* we can assume that the deaths are rather completely registered, then these death rates and the corresponding life expectancy figures provide little or no support to the view that the Hispanic population of Texas is greatly underenumerated. Any substantial underenumeration would result in excessively high death rates and excessively low life expectancy values in comparison with the true values. However, the life expectancy figures are already unexpectedly high, in view of the relative socioeconomic positions of Mexican-Americans and Anglos in Texas. It does not seem reasonable to believe that they should be higher. A uniform reduction of 10 percent in the death rates of the Spanish-surname population of Texas, corresponding to a uniform adjustment of the population by 10 percent, for example, would increase life expectation at birth from 70.2 years to about 71.9 years, bringing it to about the level for the (total) White population of Texas in 1970.

Sex Ratio Analysis

A very powerful method of analyzing census figures for coverage and reporting errors is the comparison of observed sex ratios (males per 100 females) for age groups with "expected" sex ratios or sex ratios with known characteristics. A rough set of expected sex ratios for the Hispanic population is displayed in figure 1. Also shown in figure 1 are observed sex ratios for the Spanish-origin population and observed and expected sex ratios for the White and Black-and-other-races populations in 1970.

The expected sex ratios were obtained by applying 1.05, the assumed sex ratio of births, to the sex ratios of the survival rates from the life table for the Spanish-surname population of Texas in 1970; that is, the male survival rates were divided by the female survival rates and then adjusted upward for the assumed sex ratio of births. Because the mortality level represented in this life table is lower than the actual mortality of older cohorts (since it reflects only mortality in 1970), the survival rates are probably biased upward. On the other hand, this bias is at least partially offset by a downward bias attributable to the smaller gap between the actual male and female mortality in the past. The assumed sex ratio of births was selected after an examination of sex ratios of births for various populations in Latin America and selected Hispanic populations in the United States, such as the Spanish-surname population of Texas and the Puerto Rican population of New York City.

The overall shape of the curve of actual sex ratios for age groups for the Spanish-origin population in 1970 differs only moderately from that of the expected sex ratios based on the life table. At ages 0 to 14, there appears to be a small deficit of the actual sex ratios relative to the expected sex ratios, implying a small relative deficit of males compared with females. However, comparison of sets of sex ratios in the age range 15 to 39 suggests that there is a pronounced relative deficit of males over this age range, with the sex ratios at ages 25 to 34 being almost 15 points lower than the expected values. Over the age range 40 to 59, the shape of the curve of observed sex ratios is similar to the curve of expected sex ratios, but the observed levels are a few points lower. This difference may also indicate a larger undercount of males than females. However, the level of the expected curve may be in error for various other reasons. The assumed sex ratio of births may be too high; the male-female differences in survival rates actually experienced in these and younger ages may have been greater than shown by the life tables used to compute the expected sex ratios; or the sex balance of net immigration may have been more "female" than "male" in these and younger ages.

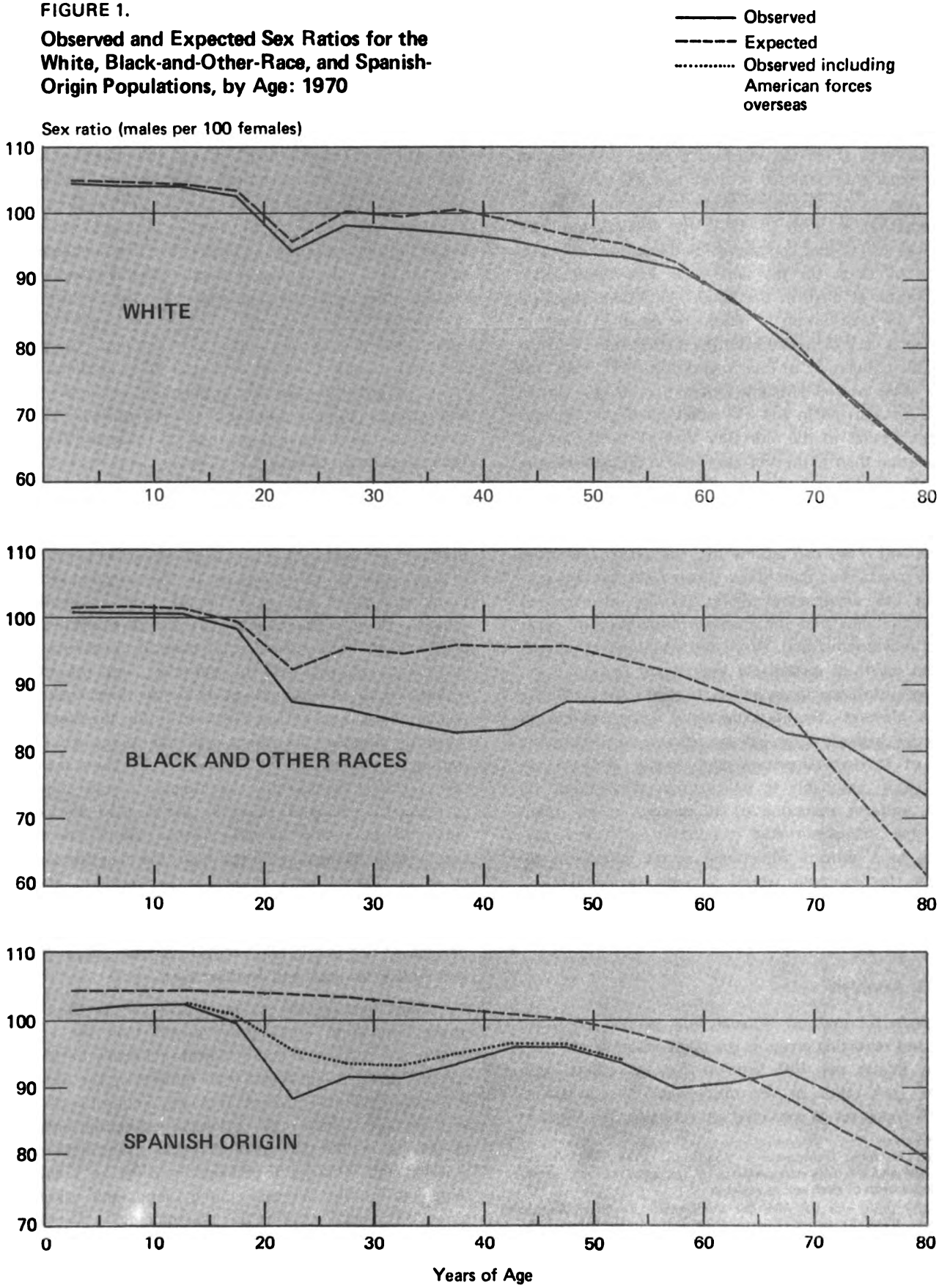
At ages over 65, the sex ratios observed for the Spanish-origin population are slightly greater than the expected values. This difference could indicate relatively better coverage of Spanish-origin males than females at the older ages. Another possibility, however, is greater emigration of women at these ages. There may have been substantial emigration to Mexico on the part of older persons who have chronic illnesses and who go to Mexico in order to spend their remaining months or years among friends and relatives "at home." The occurrence of this phenomenon is suggested by an examination of the distribution of deaths by cause among Mexican-Americans in the United States. A disproportionate share of deaths at the older ages among Mexican-Americans in Southwestern United States results from violence and acute causes; there are relatively few deaths from chronic

[42] United Nations, **Demographic Yearbook, 1973**, New York, 1974, pp. 344 and 84. Life expectancies by sex were averaged, using the total population of each sex as weights.

[43] The life table was supplied by the Division of Demographic Registry and Vital Statistics, Puerto Rico Department of Health. Cause-of-death statistics for Puerto Rico were taken from: National Center for Health Statistics, **Vital Statistics of the United States Volume I, 1970**. The distribution of deaths by cause for the U.S. population of Puerto Rican birth is represented by statistics for New York City which apply to the city's population of Puerto Rican birth. The data were supplied by the New York City Department of Vital Statistics.

FIGURE 1.

Observed and Expected Sex Ratios for the White, Black-and-Other-Race, and Spanish-Origin Populations, by Age: 1970



Note: Points are plotted at center of each 5-year age interval; last point (age 80) denotes 75 years and over.

causes (e.g., the so-called degenerative diseases).[44] To the extent that such migration is selective of women and that deaths due to violence and acute causes are more prevalent among males, the observed sex ratios of death rates at the older ages will tend to be elevated. The hypothesis that such return migration or some similar phenomenon frequently occurs is supported also by the fact that, among the various Spanish-origin populations, only the Mexican-origin group exhibits very high sex ratios at the older ages (figure 2).

The sex ratios of the populations classified by the major Hispanic identifiers in 1970 all give roughly the same indications. The curves of the sex ratios for the Spanish-heritage, Spanish-language, and Spanish-surname populations are virtually identical (figure 3). These curves have the same general form as the curve for the Spanish-origin population, although at ages over 25 the Spanish-origin population has somewhat lower sex ratios. This difference may be an indication of a greater tendency for adult females to designate themselves as being of Spanish origin. Another possible explanation for the higher sex ratios in populations identified wholly or partly by surname is the previously mentioned loss of women from the Spanish-surname population through marriage.

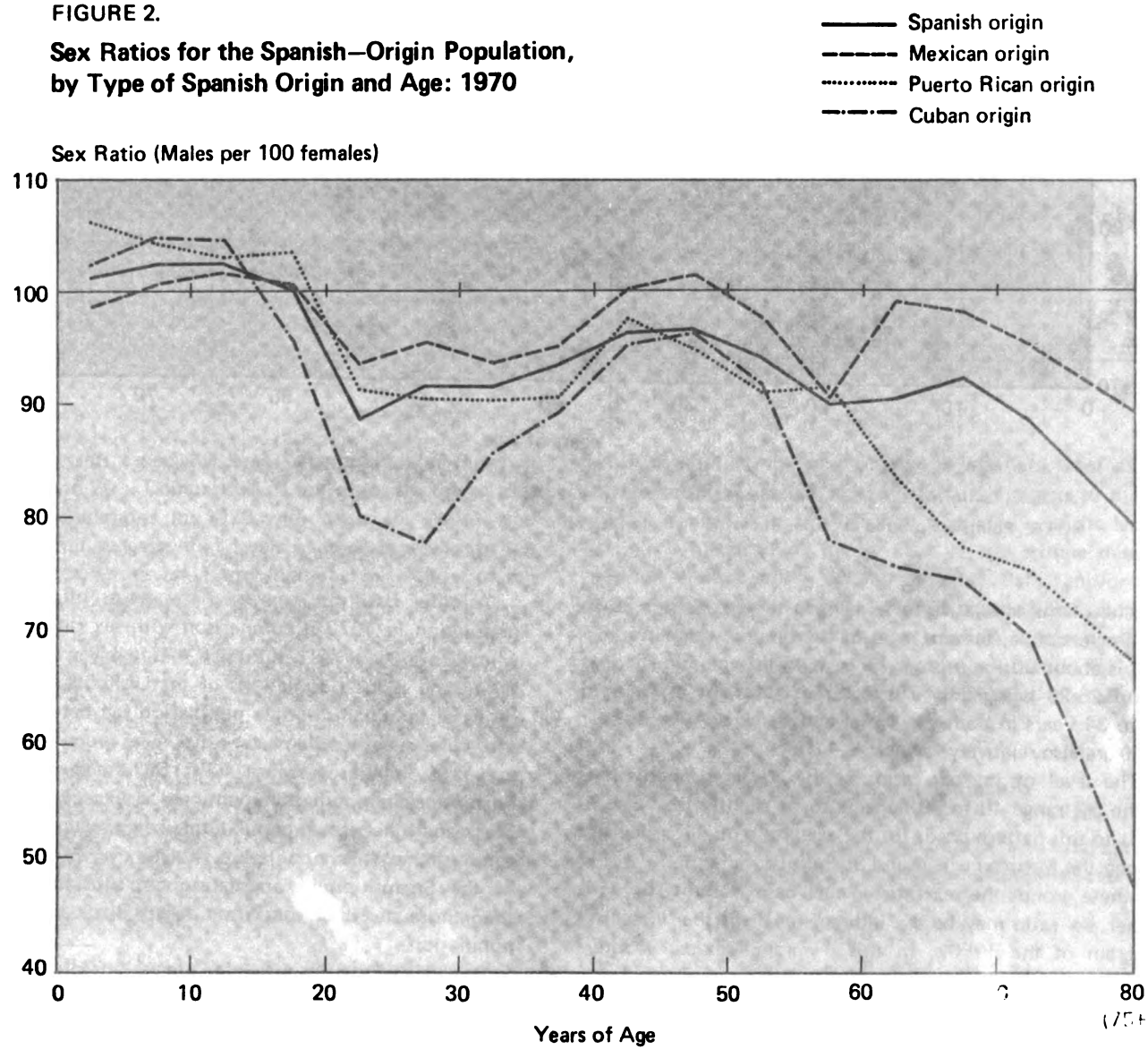
The implications of these sex ratios for estimating coverage are the same for the various identifiers. At ages under 20, the various Hispanic populations have sex ratios in 1970 roughly similar to the expected sex ratios. One exception is the sex ratio of the 1970 Mexican-origin population under 5 years of age; the figure is less than 100 and suggests a larger undercount of males relative to the undercount of females at these ages.

The sex ratios of the various Hispanic populations in 1970 imply a substantial undercount of young adult males regardless of which Hispanic identifier is used. They indicate large

[44] Benjamin S. Bradshaw and Edwin Fonner, Jr., "The Mortality of Spanish Surnamed Persons in Texas: 1969-71," in Frank D. Bean and W. Parker Frisbie (eds.), **The Demography of Racial and Ethnic Groups**, Academic Press, forthcoming.

FIGURE 2.

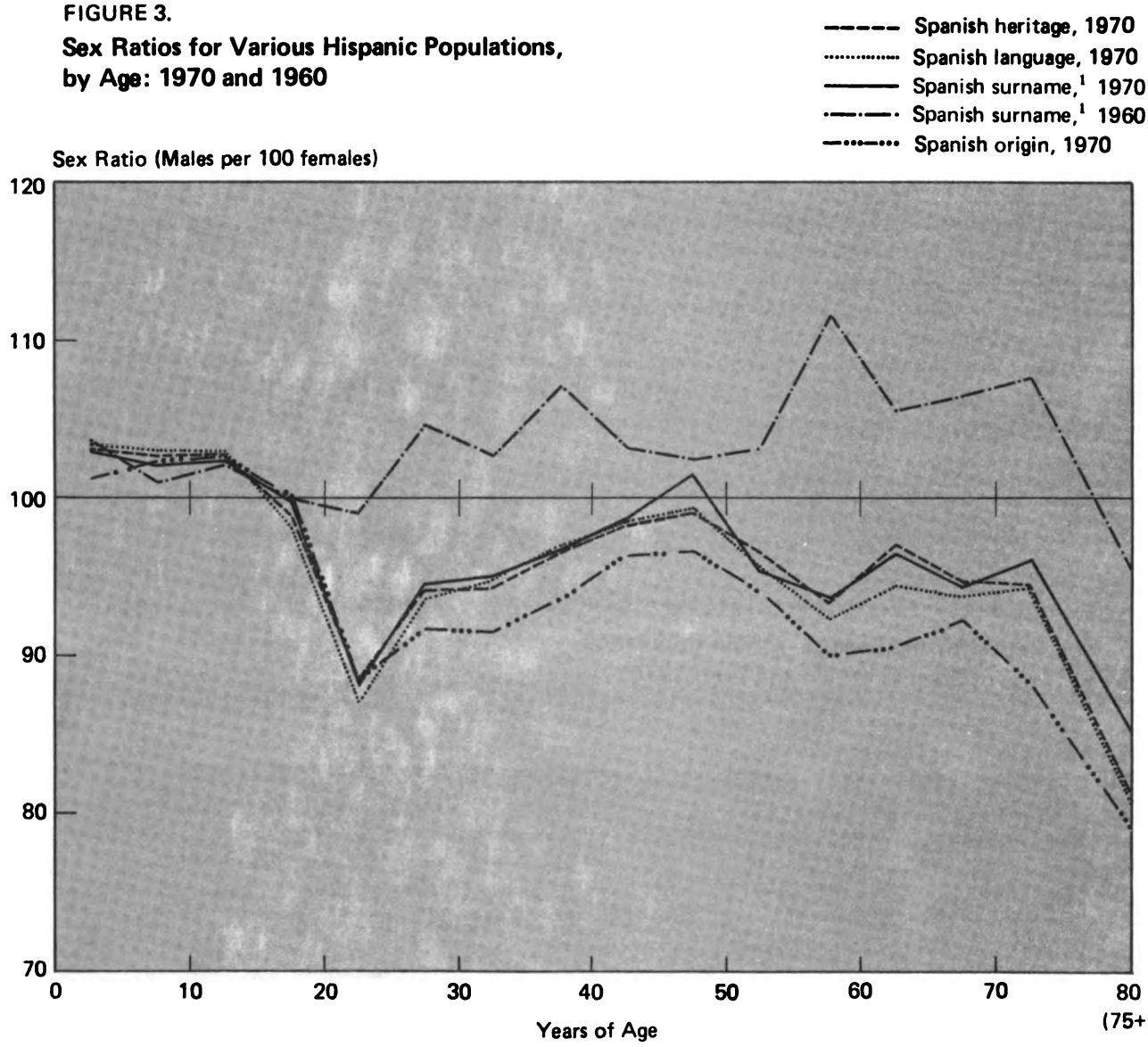
Sex Ratios for the Spanish–Origin Population, by Type of Spanish Origin and Age: 1970



Note: Points are plotted at center of each 5–year age interval.

FIGURE 3.

Sex Ratios for Various Hispanic Populations, by Age: 1970 and 1960



[1] Arizona, California, Colorado, New Mexico, and Texas only.

NOTE: Points are plotted at center of each 5-year age interval.

deficits of males relative to females in the age groups 20 to 39. For example, the sex ratio of the Spanish-origin population is about 90 for this age range with the value for the age group 25-29 being only 78. Coverage rates for males aged 25 to 34 years in both the White and Black populations for 1970 are also relatively low (table 11).

The level of the sex ratios of the Hispanic populations in the age range 40 to 59 appears to be slightly low. Exceptions to this pattern occur for the age group 45-49, especially among the Mexican-origin and Spanish-surname populations; for these groups the sex ratio in each case is about 101. This higher sex ratio may be the indirect result of the "bracero" program of the 1950's. In this program Mexican laborers were imported into the United States on a temporary basis. Since most of the laborers were male and many remained in the United States, the population sex ratio may have been significantly raised.

Another basis for analyzing sex ratios of the Hispanic population in 1970 is comparison with sex ratios of groups whose coverage levels are known reasonably well such as the White and Black populations of the United States. The sex ratios of the Spanish-origin population fall between those of the White and the Black-and-other-races populations at ages under 40. In the age range 40 to 59, the sex ratios of the Spanish-origin population continue to exceed those of the Black-and-other-races population, being roughly equal to those of the White population. At ages over 60 the sex ratios of the Spanish-origin population are substantially greater than those of either the White or the Black-and-other-races population.

Analysis of the sex ratios for the various Hispanic populations in 1970 suggests the following tendencies in census coverage. Males appear to be covered less well than females. There appear to be substantial undercounts at the young

Table 11. Preferred Estimates of the Percents of Net Undercount of the Population of the United States, by Sex, Race, and Broad Age Group: 1970 and 1960

(Base of percentages is corrected population. Minus sign (-) indicates a net overcount)

Year and age (years)	Both sexes		Male		Female	
	White	Black	White	Black	White	Black
1970						
All ages	1.9	7.7	2.5	9.9	1.4	5.5
Under 5	2.1	10.1	2.3	10.4	2.0	9.8
5 to 9	2.3	7.3	2.4	7.7	2.2	6.9
10 to 14	1.0	3.2	1.1	3.5	0.9	2.8
15 to 19	0.9	3.7	1.3	4.3	0.5	3.2
20 to 24	1.8	8.5	2.5	12.1	1.1	5.2
25 to 34	3.4	12.5	4.3	18.5	2.4	6.7
35 to 44	2.0	10.7	3.6	17.7	0.5	4.0
45 to 54	1.4	8.7	2.7	12.4	0.1	5.3
55 to 64	1.1	8.0	2.2	9.2	1.9	7.0
65 and over	1.8	1.2	1.2	-3.1	2.2	4.2
1960						
All ages	2.0	8.0	2.4	9.7	1.6	6.3
Under 5	1.5	5.8	1.9	6.6	1.1	5.1
5 to 9	1.9	4.7	2.4	5.1	1.5	4.2
10 to 14	2.0	4.4	2.5	5.0	1.5	3.9
15 to 19	3.2	10.9	3.8	12.3	2.4	9.6
20 to 24	3.4	13.9	4.3	18.4	2.4	9.5
25 to 34	2.3	12.5	3.6	18.5	1.0	6.5
35 to 44	1.0	7.6	2.2	11.5	-0.2	3.8
45 to 54	2.5	9.9	2.5	11.0	2.4	9.0
55 to 64	1.1	10.1	0.5	8.5	1.7	11.6
65 and over	2.2	-1.0	0.0	-5.8	3.5	2.8

Source: U.S. Bureau of the Census, Census of Population and Housing: 1970, Evaluation and Research Program PHC(E)-4, Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis, 1974, table 6, p. 31.

adult ages, with especially large undercounts for males in their 20's and 30's. Comparison of sex ratios for the various Hispanic populations in 1970 with those for Whites and Blacks further suggests that levels of undercoverage for the Hispanic populations are intermediate between those of the White and Black populations (see table 11).

SPECIFIC MEASURES OF NET ERRORS IN AGE-SEX DATA

The demographic techniques which have been used to derive specific measures of net errors in age-sex data from censuses include construction of an expected population from vital statistics and immigration data, comparative analysis of data from a series of previous censuses and reconstruction of the population for the date of the last census (e.g., the Coale-Zelnik and Coale-Rives methods),[45] intercensal cohort analysis, and comparison with aggregate data from administrative records. Application of such techniques to the Hispanic population in 1970 is not possible or, at the least, is seriously handicapped by the lack of appropriate data, the limited comparability of the available data, and/or their restricted temporal and geographic scope. Some examples of attempts to estimate census coverage and reporting of specific age groups for the Hispanic population by demographic techniques, particularly intercensal cohort analysis, are presented in this part of the report.

Comparison with Expected Population

Construction of an expected population from vital statistics makes use of a form of the population component estimating equation

$$\hat{P}_a^{1970} = B^{1970-(a+1)} - D_a + M_a$$

That is, the expected population aged a in 1970 is equal to the births a to a+1 years earlier minus the deaths that have

[45] Ansley J. Coale and Melvin Zelnik, **New Estimates of Fertility and Population in the United States**, Princeton University Press, 1963.
Ansley J. Coale and Norfleet W. Rives, Jr., "Statistical Reconstruction of the Black Population of the United States, 1880-1970: Estimates of True Numbers by Age and Sex, Birth Rates, and Total Fertility," **Population Index**, January 1973.

occurred to the birth cohort between the date of birth and the census date plus net migration of this birth cohort.

As has been noted, practically none of these data exist for the Hispanic population for the entire country according to any of the identifiers. Only very limited data on Hispanic births exist—births to persons of Spanish surname for about five States, Puerto Rican births for one State (New York), and corresponding data for selected cities (e.g., Los Angeles, San Antonio, New York City)—and even these data do not go back very far in time.[46] For deaths, the situation is similar. Only a limited amount of data on the international and internal migration of the Hispanic population exists even though migration is an important contributor to the growth of the Hispanic population. Since adequate data of this kind are not available, the construction of an expected population and the estimation of coverage errors for the Hispanic population in 1970 by applying this particular demographic technique cannot be satisfactorily accomplished.

Intercensal Cohort Analysis: General Considerations

Intercensal cohort analysis uses the component estimating equation to follow a cohort from one census to another

$$\hat{P}_a^{1970} = P_{a-10}^{1960} - D_a^{1960\text{-}70} + M_a^{1960\text{-}70} \quad , a \geq 10$$

Here the components $D_a^{1960\text{-}70}$ and $M_a^{1960\text{-}70}$ refer to the data for the intercensal period for the cohort age a-10 in 1960 and a in 1970. Comparison of the expected population in a given age group in 1970, $\hat{P}_a^{1970}$, with the census count for 1970 in the age group yields a measure of "bicensal relative error," i.e., a measure of the net error in the 1970 count relative to the net error in the 1960 census count for the same age *cohort.* This measure, calculated by dividing the difference between the census count and the expected population by the expected population, can be interpreted loosely as the number of percentage points by which the "coverage" of a cohort improved or worsened relative to its "coverage" in the previous census.[47]

In order to derive a measure of absolute error in the 1970 census count by means of intercensal cohort analysis, it is necessary to correct the 1960 population count, P_{a-10}^{1960}, for net census error. Since definitive estimates of the coverage of the 1960 census are not available for any of the Hispanic populations, the results of intercensal cohort analysis for these populations are limited to bicensal relative errors or illustrative calculations based on assumed levels of coverage for 1960. Furthermore, the limited data available on the mortality and migration of the Hispanic population for the 1960-70 intercensal period make any results subject to possibly serious errors.

The lack of historical comparability and the limited geographic scope of the data also preclude a meaningful application of intercensal cohort analysis on a national scale to the estimation of the coverage of the Hispanic population. Data are not available for the entire country according to any general Hispanic identifier for the last two censuses. As previously noted, only two of the Hispanic identifiers employed in 1970 were used in the previous census—Spanish surname (in five Southwestern States) and birth or parentage in a country of Spanish language (Cuba, Puerto Rico, Mexico, etc.). However, intercensal cohort analysis is not feasible for the Spanish-surname population. Counts of the Spanish-surname population in the 1960 and 1970 censuses were produced only for the five Southwestern States. This geographic restriction, combined with a lack of data on the migration of persons of Spanish surname to and from the five States, prevents any possibility of deriving adequate coverage estimates for this population. In fact, intercensal cohort analysis for the Spanish-surname population would probably produce better estimates of intercensal net migration than of coverage.

Intercensal cohort analysis can be meaningfully applied to certain populations of Hispanic birth or parentage for cohorts alive in 1960 (i.e., aged 10 years and over in 1970). At best, such an analysis can tell us only about the relative consistency of coverage and reporting in the two censuses for the restricted population groups involved. The estimates would be affected by any errors in the allowances for intercensal mortality and net migration. In applying intercensal cohort analysis when there is no need to allow for net migration or when satisfactory data on net migration are available, relatively more accurate estimates of coverage and reporting errors may be derived. Accordingly, illustrative estimates of the coverage of the population of Puerto Rican birth or parentage, the population of Cuban birth, and the native population of Mexican parentage will be presented. The specific circumstances that permit the preparation of these estimates are discussed in subsequent sections.

Intercensal Cohort Analysis: Population of Puerto Rican Birth or Parentage

Data on the population of Puerto Rican birth or parentage in the United States are available for each of the components in the intercensal estimating equation. Detailed tabulations for sex and age categories are available from both the 1960 and 1970 censuses for the population of Puerto Rican birth or parentage.[48] Life tables applicable to various time periods

[46] In 1978 about 18 States started to collect birth statistics identifying births of the Spanish-origin population. See the subsequent section, "Prospects for Measuring the Coverage of the Hispanic Population."

[47] Bicensal relative error algebraically represents the change in the amount of undercount between the two censuses divided by the excess of the 1970 true population (unknown) over the amount of undercount in 1960 (also unknown). As such, the bicensal relative error is an upper limit to the percentage point change in net census error for an age cohort under most conditions. Another similar interpretation of the bicensal relative error is that it represents the percent net "coverage" error in the second census on the assumption that the first census is a perfect count. If the amount of intercensal change exceeds the amount of undercount in the earlier census, as it will for most populations and for age cohorts subject to in-migration, then the bicensal relative error will be greater than the change in undercount rates.

[48] U.S. Bureau of the Census, Census of Population: 1970, Subject Reports, PC(2)-1E, **Puerto Ricans in the United States,** 1973, and U.S. Bureau of the Census, Census of Population: 1960, Subject Reports, PC(2)-1D. **Puerto Ricans in the United States,** 1963.

in the decade have been developed for the population of Puerto Rico and were assumed to apply to the United States population of Puerto Rican birth or parentage for measuring the mortality component of the population.[49] Estimates of net migration, usually the most elusive component, were developed from the monthly totals of passenger traffic entering and leaving Puerto Rico compiled by the Puerto Rico Planning Board. Estimates of the corresponding age-sex distribution of net migration are more problematical than other elements of the intercensal cohort analysis but can be derived from several small surveys conducted in Puerto Rico and from the 1960 and 1970 censuses of both the United States and Puerto Rico. Subsequent sections describe the specific procedures and assumptions employed to produce estimates of bicensal relative coverage of the population of Puerto Rican birth or parentage.

A basic step in deriving the expected population of Puerto Rican birth or parentage in 1970 is to calculate the survivors in 1970 of the population of Puerto Rican birth or parentage in the United States in 1960. To determine the size of this population for each age-sex group in 1970, 5-year (age-interval) life table survival rates from the Puerto Rican life tables for 1959-61 and 1967-69 were applied to the enumerated population of Puerto Rican birth or parentage in 1960. (See table 12.) The levels of life expectation at birth in the Puerto Rican life tables—67.1 years and 68.2 years for males in 1959-61 and 1967-69, and 71.9 years and 73.8 years for females in 1959-61 and 1967-69, respectively—are similar to the levels in the life tables for the Spanish-surname population of Texas and the Puerto Rican population of New York City. Thus, it was deemed appropriate to apply the Puerto Rican life tables to Puerto Ricans residing in the United States.

The migration component for the intercensal period was not derived in such a straightforward manner. The Puerto Rico Planning Board supplies the Bureau of the Census with monthly totals of arrivals and departures by air and water for Puerto Rico. These figures are adjusted slightly to remove seasonal fluctuations and the difference between them gives the net migration from Puerto Rico to the United States. However, information on the age and sex distribution of the migrants is not available from the same source. The age and sex distribution of migrants to Puerto Rico was derived from the Labor Force Survey conducted by the Puerto Rico Planning Board covering the years 1965 through 1967. The age and sex distribution of migrants from Puerto Rico to the United States came from the Health Survey (a subsample of the Labor Force Survey) for the years 1963 and 1965 through 1967. These two age-sex distributions were combined to give a distribution for "net migrants" which was assumed to apply to each year of the decade.

The annual net migration figures for the 1960-1970 decade from Puerto Rico to the United States were recompiled into age cohorts for the two periods April 1, 1960 to March 31, 1965 and April 1, 1965 to March 31, 1970. The migrants for the first half of the decade were "survived" to April 1, 1970, the census date, by applying 2½-year survival rates from the 1959-61 Puerto Rican life tables and then 5-year survival rates from the 1967-69 Puerto Rican life tables. For migrants entering in 1965-70, 2½-year survival rates from the 1967-69 life tables were used. The resulting estimates of surviving migrants were added to survivors of the 1960 census population to derive the expected population in 1970. Comparison of this expected population with the 1970 census counts of the population of Puerto Rican birth or parentage then gives a measure of the improvement or deterioration in census coverage since 1960 for each cohort. The results are shown in table 12.

Intercensal cohort analysis indicates a slight improvement, 0.9 percent at most, in the coverage of the population of Puerto Rican birth or parentage between the 1960 and 1970 censuses. Although the estimate of coverage for the total population is plausible, many of the estimates for specific age and sex groups are clearly implausible and the possibility of errors in the estimation procedure is indicated. The coverage rates for each sex, a 6.2 percent decline in coverage for males and an 8.7 percent improvement for females, taken separately, are each within the realm of possibility. However, there is no good explanation why the coverage of one sex would improve so dramatically while the other is experiencing an almost equally dramatic decline.

The results in table 12 indicate that the coverage of the total cohort aged 20-29 in 1970 (10-19 in 1960) was 24 percent worse, and the male cohort 34 percent worse, in 1970. It is extremely unlikely that coverage decreased by such a tremendous percentage even for this age cohort, which usually has greater decreases in coverage than other age cohorts. (See table 11 for coverage rates in 1970 and 1960 for Whites and Blacks.) For females over age 30 and males over age 40 in 1970 coverage seems to have improved between 1960 and 1970. However, the magnitudes of the improvements, from 25 to 99 percent for males and 28 to 94 percent for females, are much too large to be plausible. Coverage of the population of Puerto Rican birth or parentage in the 1960 census could not have been so poor as to leave room for improvements of this magnitude.

The pattern of large relative overcounts in some age-sex groups and large relative undercounts in others, combined with a plausible coverage level for the total population, indicates the possibility of misallocation of one or more components by age and sex. The most likely source of this type of error is the age-sex distribution of the net migrants. Although the total number of migrants was determined from a complete count of traffic to and from Puerto Rico, the age and sex distribution of these migrants was determined from small sample surveys covering only a part of the decade. Consequently, for re-estimation of the coverage of Puerto Ricans, the total number of migrants for each half of the decade, approximately 46,000 from 1960 to 1965 and 112,000 from 1965 to 1970, was accepted as accurate, but the age-sex distribution of the migrants derived from the Puerto Rican sample surveys was rejected as unsatisfactory and an alternative source was investigated.

[49] Life tables for 1959-61, 1967-69, and 1969-71, for each sex, were supplied by the Division of Demographic Registry and Vital Statistics, Department of Health, Puerto Rico.

Table 12. Estimates of Bicensal Relative Error for the Population of Puerto Rican Birth or Parentage, Using Sample Survey Data for the Age-Sex Distribution of Migrants, by Sex: 1960-1970

(Populations in thousands. Percentages and totals computed from unrounded figures)

Age (years) and sex	Census population, 1960[1] (1)	Survivors of 1960 census, 1970[2] (2)	Net movement from Puerto Rico, 1960-1970[3] (3)	Survivors of net movement from Puerto Rico, 1970[3,4] (4)	Expected population, 1970 (2)+(4)= (5)	Census population, 1970[1] (6)	Percent relative error[5] [(6)-(5)]÷(5)= (7)
In 1960 In 1970							
TOTAL							
All ages.....10 and over.....	893.0	859.0	+141.3	+146.2	1,005.1	1,014.2	+0.9
Under 10.....10 to 19.....	255.1	252.4	+66.3	+66.0	318.5	308.2	-3.2
10 to 19.....20 to 29.....	165.9	164.2	+175.3	+174.2	338.5	257.8	-23.8
20 to 29.....30 to 39.....	180.5	176.9	-0.2	-0.1	176.8	194.9	+10.2
30 to 39.....40 to 49.....	143.3	138.3	-38.6	-37.6	100.6	128.2	+27.4
40 to 54.....50 to 64.....	100.5	93.1	-42.5	-40.3	52.8	91.4	+73.1
55 and over.....65 and over.....	47.6	33.9	-19.0	-16.0	17.9	33.8	+88.6
MALE							
All ages.....10 and over.....	446.9	427.7	+94.9	+97.4	525.1	492.6	-6.2
Under 10.....10 to 19.....	129.2	127.7	+32.7	+32.6	160.3	156.6	-2.3
10 to 19.....20 to 29.....	81.7	80.6	+108.0	+107.2	187.8	123.0	-34.5
20 to 29.....30 to 39.....	91.0	88.6	+9.4	+9.4	98.0	94.0	-4.1
30 to 39.....40 to 49.....	73.7	70.5	-21.2	-20.6	50.0	62.2	+24.6
40 to 54.....50 to 64.....	50.5	45.9	-25.8	-24.3	21.6	43.0	+99.2
55 and over.....65 and over.....	20.8	14.5	-8.3	-6.9	7.6	13.8	+81.7
FEMALE							
All ages.....10 and over.....	446.2	431.2	+46.4	+48.8	480.0	521.6	+8.7
Under 10.....10 to 19.....	125.9	124.8	+33.5	+33.5	158.2	151.6	-4.2
10 to 19.....20 to 29.....	84.3	83.7	+67.3	+67.0	150.7	134.8	-10.6
20 to 29.....30 to 39.....	89.5	88.3	-9.6	-9.5	78.8	100.8	+27.9
30 to 39.....40 to 49.....	69.7	67.8	-17.4	-17.1	50.7	66.0	+30.3
40 to 54.....50 to 64.....	50.0	47.3	-16.7	-16.0	31.2	48.4	+55.1
55 and over.....65 and over.....	26.8	19.4	-10.6	-9.1	10.3	20.0	+93.6

[1]Males adjusted to include Armed Forces overseas.
[2]Derived with 1959-61 and 1967-69 life tables for Puerto Rico (by sex).
[3]A plus sign denotes net movement into the United States; a minus sign denotes net movement into Puerto Rico.
[4]Derived from 1960-65 and 1965-70 migration figures and 1959-61 and 1967-69 life tables for Puerto Rico (by sex).
[5]A plus sign denotes a relative net overcount in the 1970 census as compared with the 1960 census; a minus sign denotes a relative net undercount. Base of percent is expected population.

Sources: See text.

Basic sources of information on migration between Puerto Rico and the United States are the replies to the questions on place of birth and residence 5 years ago in the 1960 and 1970 censuses of the United States and Puerto Rico. Data on residence in Puerto Rico in 1965 from the 1970 census for age-sex groups 15 years and over of the population of Puerto Rican birth and parentage were adjusted to include a proportion of persons who had moved but did not report their residence.[50] These data, given only in broad age groups, were subdivided into 5-year groups on the basis of the distribution of the corresponding age-sex groups for the population of Puerto Rican birth or parentage residing in New York, New Jersey, and Pennsylvania. To estimate migrants from Puerto Rico from 1965 to 1970 at ages under 15, the proportion of each Puerto Rican birth cohort migrating to the United States in a 5-year period was first calculated by combining the population residing in the United States but born in Puerto Rico with the population residing and born in Puerto Rico, as shown by the 1960 and 1970 censuses. Successive differences in lifetime proportions migrating for age cohorts in the two censuses then gave an estimate of the proportion of the birth cohort migrating to the United States from 1965 to 1970. The result of applying these procedures was a complete age-sex distribution of migrants from Puerto Rico to the United States during the 1965-70 period.

The procedure for estimating migration from the United States to Puerto Rico was similar. Data for 5-year age-sex groups over age 15 on residence in the United States in 1965 from the 1970 census for Puerto Rico were adjusted *pro rata* to include a proportion of those who had moved but failed to report their residence in 1965.[51] For ages under 15, the calculation paralleled that for migrants from Puerto Rico except that the cohorts born in the United States consisted of the sum of persons of Puerto Rican parentage residing in the United States in 1970 and persons born in the United States and residing in Puerto Rico in 1970. These procedures yielded a complete age-sex distribution of persons migrating from the United States to Puerto Rico during the 1965-70 period.

The differences between the age-sex distribution of migrants from Puerto Rico and the age-sex distribution of migrants to Puerto Rico from 1965 to 1970 yielded a distribution by age and sex for net migration to the United

[50]U.S. Bureau of the Census, Census of Population: 1970, Subject Reports, PC(2)-1E, **Puerto Ricans in the United States**, 1973, table 5, p. 39.

[51]U.S.Bureau of the Census, Census of Population: 1970, Volume 1, **Characteristics of the Population, Part 53, Puerto Rico**, PC(1)-53, 1973, table 113, p. 53-623.

States from Puerto Rico. This age-sex distribution was then assumed to apply both to net migrants for the 1965-70 period in cohorts aged 5 years and over in 1970 and for the 1960-65 period in cohorts aged 10 years and over in 1970. The distribution was adjusted by the plus-minus proportionate adjustment method[52] for each 5-year period to the previously established totals derived from the Puerto Rico Planning Board data. The net migrants in each 5-year period were then "survived" to 1970 by use of the same life tables as in the previous estimate and then added to survivors of the 1960 census counts, to yield the estimates of the expected population of Puerto Rican birth or parentage shown in table 13.[53]

The coverage estimates for the population of Puerto Rican birth or parentage in the 1970 census relative to the 1960 census which result from use of the revised migration estimates are clearly much more plausible than those originally derived. These estimates indicate an improvement in coverage of 4.9 percent for the total population 10 years and over in 1970 (3.9 percent for males and 5.8 percent for females). The largest improvements are for the male and female cohorts aged 30-39 in 1970 (20-29 in 1960), 16 percent for males and 22 percent for females. The ages 20-29 are usually among the most poorly covered (table 11), so that this amount of improvement is consistent with an overall improvement in coverage and the improvement attributable to the aging of the cohort. Other large improvements in coverage occurred for the male cohorts aged 65 and over in 1970 and female cohorts aged 50 and over in 1970. These large coverage increases relative to the 1960 census are consistent with the inferences drawn from the analysis of census survival ratios regarding the existence of possibly large coverage errors in 1960 in these cohorts. Relative coverage errors in the other cohorts are generally small and appear plausible.

Again, it must be stressed that the coverage estimates for the population of Puerto Rican birth or parentage shown in table 13 represent bicensal relative errors, i.e., changes in levels of census coverage between 1960 and 1970. In order to derive measures of the absolute level of census coverage in 1970 for the population of Puerto Rican birth or parentage using intercensal cohort analysis, estimates of coverage in the 1960 census are required. Since such estimates are not available, definitive estimates of coverage for the population of Puerto Rican birth or parentage in 1970 cannot be developed.

[52] **The Methods and Materials of Demography**, op. cit., pp. 705-6.

[53] Discrepancies in the migration totals shown in tables 12 and 13 are the result of differences in the number of migrants allocated to the cohorts born between 1960 and 1970.

Table 13. Estimates of Bicensal Relative Error for the Population of Puerto Rican Birth or Parentage, Using Census Data on Migration, by Sex: 1960-1970

(Populations in thousands. Percentages and totals computed from unrounded figures)

Age (years) and sex		Census population, 1960[1] (1)	Survivors of 1960 census, 1970[2] (2)	Adjusted net migration from Puerto Rico, 1960 to 1970[3 4] (3)	Surviving net migrants from Puerto Rico, 1970[4 5] (4)	Expected population, 1970[1] (5)	Census population, 1970[1] (6)	Percent relative error[6] [(6)-(5)]÷(5)= (7)
In 1960	In 1970							
TOTAL								
All ages	10 and over	893.0	859.0	+107.7	+108.3	967.3	1,014.2	+4.9
Under 10	10 to 19	255.1	252.4	+58.8	+58.6	311.1	308.2	-0.9
10 to 19	20 to 29	165.9	164.2	+85.7	+85.3	249.6	257.8	+3.3
20 to 29	30 to 39	180.5	176.9	-13.1	-13.0	164.0	194.9	+18.8
30 to 39	40 to 49	143.3	138.3	-11.6	-11.3	126.9	128.2	+1.0
40 to 54	50 to 64	100.5	93.1	-7.6	-7.3	85.8	91.4	+6.5
55 and over	65 and over	47.6	33.9	-4.6	-4.1	29.9	33.8	+13.1
MALE								
All ages	10 and over	446.9	427.7	+46.3	+46.5	474.2	492.6	+3.9
Under 10	10 to 19	129.2	127.7	+27.7	+27.5	155.2	156.6	+0.9
10 to 19	20 to 29	81.7	80.6	+39.1	+38.8	119.4	123.0	+3.0
20 to 29	30 to 39	91.0	88.6	-7.3	-7.3	81.4	94.0	+15.5
30 to 39	40 to 49	73.7	70.5	-7.1	-7.0	63.5	62.2	-2.1
40 to 54	50 to 64	50.5	45.9	-4.2	-4.0	41.9	43.0	+2.6
55 and over	65 and over	20.8	14.5	-1.8	-1.7	12.8	13.8	+7.4
FEMALE								
All ages	10 and over	446.2	431.2	+61.5	+61.8	493.1	521.6	+5.8
Under 10	10 to 19	125.9	124.8	+31.2	+31.1	155.9	151.6	-2.8
10 to 19	20 to 29	84.3	83.7	+46.7	+46.5	130.2	134.8	+3.5
20 to 29	30 to 39	89.5	88.3	-5.7	-5.7	82.6	100.8	+22.1
30 to 39	40 to 49	69.7	67.8	-4.4	-4.4	63.4	66.0	+4.2
40 to 54	50 to 64	50.0	47.3	-3.4	-3.3	44.0	48.4	+10.1
55 and over	65 and over	26.8	19.4	-2.8	-2.4	17.0	20.0	+17.3

[1] Males adjusted to include Armed Forces overseas.

[2] Derived with 1959-61 and 1967-69 life tables for Puerto Rico.

[3] Age distribution of net migrants to Puerto Rico, 1965-70, as estimated from 1970 census adjusted to total net migration from Puerto Rican passenger traffic for 1960-65 and 1965-70.

[4] A plus sign denotes net movement into the United States; a minus sign denotes net movement into Puerto Rico.

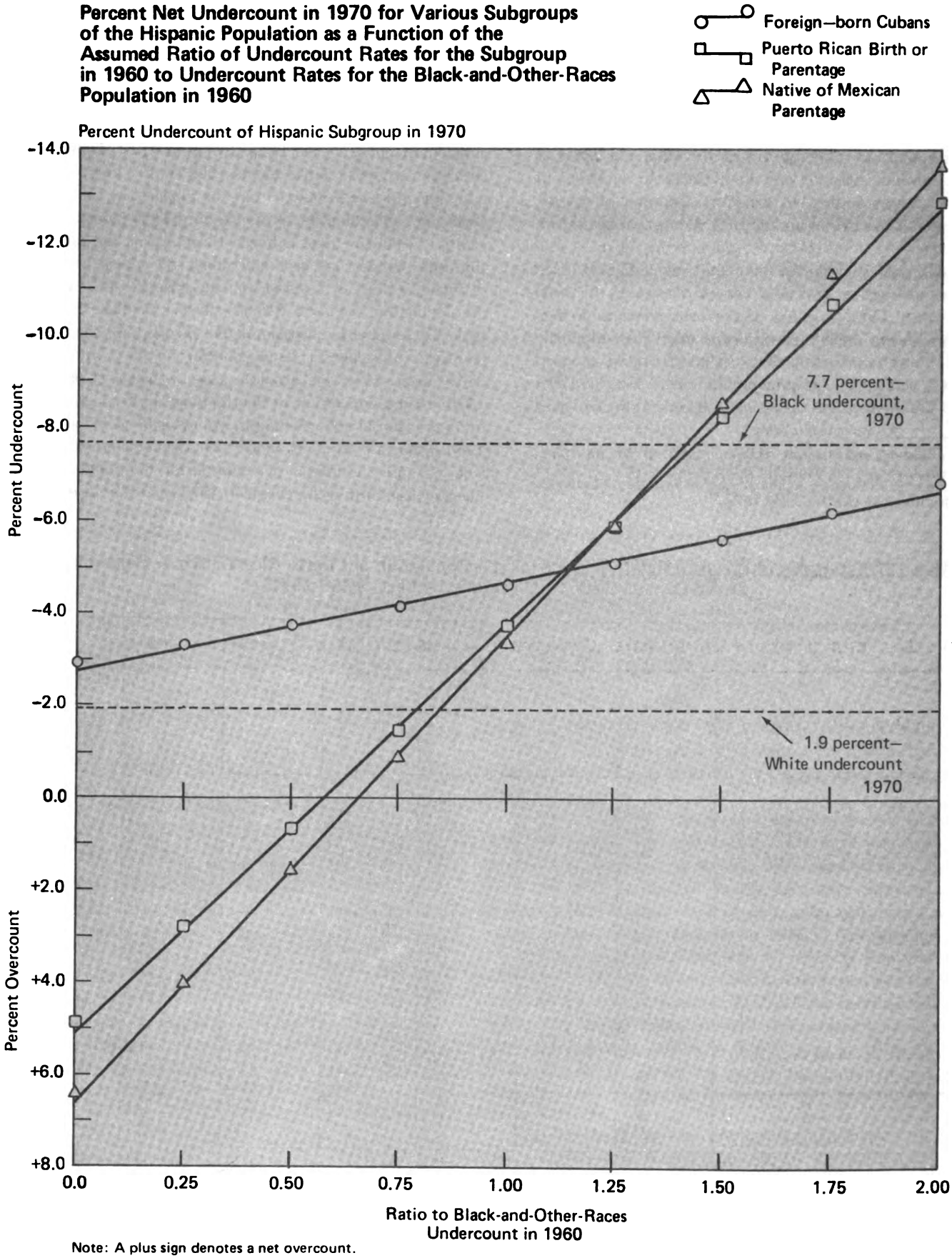
[5] Derived from 1960-65 and 1965-70 net migration figures and 1959-61 and 1967-69 Puerto Rican life tables (by sex).

[6] A plus sign denotes a relative net overcount in the 1970 census as compared with the 1960 census; a minus sign denotes a relative net undercount. Base of percent is expected population.

Sources: See text.

FIGURE 4.

Percent Net Undercount in 1970 for Various Subgroups of the Hispanic Population as a Function of the Assumed Ratio of Undercount Rates for the Subgroup in 1960 to Undercount Rates for the Black-and-Other-Races Population in 1960



Note: A plus sign denotes a net overcount.

Illustrative estimates of census coverage for the population of Puerto Rican birth or parentage in 1970 can be developed, however, by correcting the 1960 census counts for this population with an assumed set of coverage rates. For this purpose it was arbitrarily assumed that the coverage rates for the population of Puerto Rican birth or parentage in 1960 were the same (for age and sex categories) as those for the Black-and-other-races population in 1960.[54] With this assumption (which implies an overall undercount of 9.0 percent in 1960), the undercount of the population of Puerto Rican birth or parentage in 1970 would be 3.7 percent. This analysis can be carried further to determine the sensitivity of the 1970 undercount rates to various assumptions regarding the 1960 undercount rates. Figure 4 shows the 1970 undercount rates for the population of Puerto Rican birth or parentage which would result if various multiples of the 1960 undercount rates (for age and sex categories) for the Black-and-other-races population are assumed to apply to the 1960 population of Puerto Rican birth or parentage. An undercount greater than 0.78 times that of Blacks-and-other-races in 1960, or 7.1 percent, would result in a greater undercount of the population of Puerto Rican birth or parentage in 1970 than that of Whites in 1970 (1.9 percent). An undercount of less than 1.42 times that of Blacks-and-other-races in 1960, 12.9 percent, would yield an undercount in 1970 of less than the 7.7 percent undercount of Blacks. Thus, for a fairly wide range of assumed undercount rates in 1960, the undercount rate for the population of Puerto Rican birth or parentage in 1970 would be intermediate between the rates for Whites and Blacks in 1970.

[54] U.S. Bureau of the Census, Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis, PHC(E)-4, op. cit., table 5, set D.

Intercensal Cohort Analysis: Foreign-born Cuban Population

The availability of 1960 census data on the population of Cuban birth and Immigration and Naturalization Service (INS) data on immigration into the United States from Cuba between 1960 and 1970 suggests the possibility of calculating some rough estimates of net census errors for the population of Cuban birth in 1970. This group accounted for over 80 percent of the population of Cuban origin in the United States in 1970, so that the estimates can be thought of as applying approximately to the entire Cuban population of the United States. Furthermore, because of the large volume of immigration from Cuba during the 1960's, only about one-sixth of the 1970 population of Cuban birth lived in the United States in 1960. Because most of the estimated population is based on data on immigration during the decade, the component of the bicensal relative coverage error arising from error in the 1960 census data is small.

To estimate the expected population of foreign-born Cubans in 1970, we first estimate the segment that is the survivors of the foreign-born Cubans enumerated in 1960. The 1960 census provides only the total number of foreign-

Table 14. Immigrants of Cuban Birth, by Type of Admission, for Each Year, 1959-60 to 1969-70

Year (Ending June 30 of year shown)	Immigrants admitted[1] (1)	Parolees (refugees) admitted (2)	Parolees adjusting status (3)	Total entrants (4)
Total, 1960-70...	265,052	310,517	[2]140,582	[3]439,746
1960.....	8,283	(X)	(X)	8,283
1961.....	14,287	3,900	(X)	18,187
1962.....	16,254	58,630	(X)	74,884
1963.....	10,587	34,537	(X)	45,124
1964.....	15,808	5,390	(X)	21,198
1965.....	19,760	2,322	(X)	22,082
1966.....	17,355	32,542	(X)	49,897
1967.....	33,321	44,963	25,752	52,532
1968.....	99,312	45,136	91,520	52,928
1969.....	13,751	41,751	6,343	49,159
1970.....	16,334	41,346	12,208	45,472

X Not applicable.
[1] Includes parolees adjusting status after 1967.
[2] Cumulative total through June 30, 1970 as published in 1970 Annual Report of the Immigration and Naturalization Service, page 9. Annual figures shown add to 135,823.
[3] Sum of annual figures shown.

Sources: Columns (1)-(3)--United States Immigration and Naturalization Service, Annual Report, 1960 through 1970, especially table 9.
Column (4)--See text for description of derivation procedure.

born Cubans of each sex without any age detail. To determine the survivors of this group for each sex in 1970, the 1960 census totals for males and females were distributed by age according to the age distribution of persons born in Puerto Rico shown by the 1960 census and then "aged" to 1970 with survival rates from the 1959-61 life table for the White population of Florida.[55] The resulting totals of 37,400 males and 37,900 females were then redistributed by age according to the 1970 census age distribution of foreign-born Cubans who entered the United States prior to 1960. The effect of errors in the assumptions underlying these calculations would be small because, as noted earlier, most foreign-born Cubans in the United States in 1970 entered after 1960. However, the likely effect of the procedure is to increase very slightly the level of the estimated net undercounts. Cuban immigrants tend to be older than Puerto Rican immigrants so that, if the true age distribution in 1960 were known, fewer survivors in 1970 would have been estimated from the 1960 population.

The remaining foreign-born Cubans in the United States in 1970 represent survivors of three categories of Cuban arrivals between 1960 and 1970: immigrants with permanent status, parolees (refugees) with nonpermanent status, and parolees who had been in the United States for a minimum of 2 years and whose nonpermanent status was subsequently adjusted to permanent status under the 1966 Cuban Refugee Act. (See table 14 for data on immigrants.) Data for each group had to be treated differently to arrive at age distributions for them because of the different amount of information available for each group. INS Annual Reports provide age distributions in 10-year groups for each sex for immigrants with permanent status; these data were then subdivided into 5-year groups by polynomial (cubic) interpolation. For parolees (refugees), or immigrants with nonpermanent status, the INS reports provide only total annual figures for 1961 to 1970. Parolees who entered between 1960 and 1966 were assumed to have the same age and sex distribution as immigrants with permanent status who entered during the same year. Parolees who entered between 1967 and 1970 were distributed by age and sex according to the age-sex distribution of parolees who had registered with Cuban refugee centers in Florida in the late 1960's.

The remaining group of foreign-born Cubans, parolees who have been accepted for permanent residence, are included in the age and sex data on permanent immigrants for the years 1967 through 1970 tabulated by the Immigration and Naturalization Service. Only the total number of persons whose status was adjusted, according to year of arrival, is available. Since these refugees have been included in the age and sex data for permanent migrants in the year of adjustment of status and since they should be added in the year of arrival, they must be excluded from the INS age-sex data to avoid double counting. The parolees whose status was adjusted were assumed to have the same age and sex distribution in each year as the permanent Cuban immigrants. The number of Cuban immigrants for each age-sex group in each year between 1960 and 1970 was finally arrived at by adding the estimates for parolees with nonpermanent status for age-sex groups to the statistics for immigrants with permanent status and subtracting the estimates of parolees whose status was adjusted.

The expected number of survivors in 1970 of Cuban immigrants arriving between 1960 and 1970 was approximated by applying life table survival rates to the annual data on immigrants. The life tables used were those for the Spanish-surname population of Texas in 1970 previously described.[56] The survivors in 1970 were then recombined into conventional 5-year age groups. Some Cubans (immigrants and parolees) enter the greater United States through Puerto Rico and a small number of these remain in Puerto Rico. The latter persons must be excluded from the foreign-born Cuban population in the United States in 1970. The number of foreign-born Cubans who remained in Puerto Rico was assumed to be the number of aliens registering with the Immigration and Naturalization Service in 1970 (20,665).[57] Since no other information is available on these persons, it was arbitrarily assumed that one-quarter of them entered the greater United States during the 1960-65 period and the remainder entered during the 1965-70 period. These persons were then assigned to age and sex groups in proportion to all Cuban immigrants arriving during the appropriate time period. The expected foreign-born Cuban population then consisted of the survivors of foreign-born Cubans enumerated in 1960 plus the survivors of Cuban immigrants arriving between 1960 and 1970 less the number of aliens from Cuba registering with the Immigration and Naturalization Service in 1970 in Puerto Rico (table 15). It was assumed that there was no emigration of foreign-born Cubans.

Comparison of the expected number of foreign-born Cubans with the census figures appears to imply an estimated net undercount of about 3 percent for both sexes, about 4 percent for males, and about 3 percent for females. These figures must be treated as rough approximations at best. Moreover, because of the many broad assumptions employed in estimating the expected age and sex distributions, the estimates of net undercounts according to age, sex, and period of immigration cannot be treated as even approximate indications of the extent of error for these categories.

Several anomalous results in this set of estimates of the coverage of Cubans in the 1970 census point to specific

[55] The majority of Cubans in the United States in 1960 were White and lived in Florida. Many were probably long-time residents of the United States. See the next footnote also.

[56] The life tables for the Spanish-surname population of Texas in 1970 were the only recent life tables available representing the experience of a Hispanic population in the United States. Since most of the Cuban immigrants are young or middle-aged adults and since survival rates for these age groups are high, the results are practically insensitive to any reasonable choice of survival rates. Expectation of life at birth for alternative choices of life tables are:

	Male	Female
Texas, Spanish-surname, 1970-71	68.1	73.4
Florida, White, 1959-61	67.9	75.7
Florida, White, 1969-71	68.2	76.4
Puerto Rico, 1959-61	67.1	71.9
Puerto Rico, 1969-71	69.0	75.2

[57] U.S. Immigration and Naturalization Service, **1970 Annual Report**, Table 35, p. 105. The assumed figure includes 81 aliens of Cuban origin registering in the Virgin Islands.

Table 15. Illustrative Estimates of the Percents of Net Census Error for the Population of Cuban Birth, by Sex and Age and by Sex and Year of Entry into the United States: 1970

(Population in thousands. Base of percent is expected population. A plus sign denotes a net overcount in the census; a minus sign denotes a net undercount. Numbers rounded independently. Percentages computed from unrounded figures)

Age and year of immigration	Total			Male			Female		
	Census	Expected	Percent net error	Census	Expected	Percent net error	Census	Expected	Percent net error
AGE									
All ages, total	446.0	459.4	-2.9	208.7	218.4	-4.4	237.3	241.1	-1.6
Less than 15 years	76.4	74.9	+2.0	38.6	38.5	+0.3	37.8	36.4	+3.7
15 to 34 years	135.8	135.3	+0.4	62.2	64.3	-3.3	73.6	71.0	+3.7
35 to 54 years	157.7	173.7	-9.3	76.1	83.3	-8.6	81.5	90.4	-9.8
55 to 64 years	43.8	44.5	-1.6	19.0	19.7	-3.8	24.8	24.8	+0.2
65 years and over	32.3	30.9	+4.3	12.7	12.5	+1.9	19.5	18.4	+6.0
YEAR OF IMMIGRATION									
Prior to 1960	74.1	75.3	-1.6	35.9	37.4	-4.2	38.2	37.9	+0.9
1960 to 1964	167.9	161.8	+3.8	81.7	83.1	-1.7	86.2	78.7	+9.5
1965 to 1970	204.0	222.3	-8.2	91.1	97.8	-6.9	112.9	124.4	-9.3

Source of census figures: U.S. Bureau of the Census, Census of Population: 1970, Subject Reports, PC(2)-1A, National Origin and Language, 1973, table 17, p. 462. Census figures have been adjusted pro rata by age for year of entry not reported.

types of errors in the adjustment procedures. According to the estimates, children under 15 years of age were overcounted. In the light of our experience that pre-teenage children are usually undercounted, often substantially, the results suggest that too few parolees were allocated to this age group or too many "adjusted" parolees were removed from it.[58] A further difficulty with the estimation procedure is apparent in the much larger undercounts for ages 35 to 54 than for ages 15 to 34; from our experience with other populations, we would expect the relative levels of coverage for these two groups to be reversed or at least more similar. Again, this result suggests that too few parolees were assigned to the younger adult ages (for the 1960-66 period, especially) at the expense of the middle age groups or, alternatively, too many parolees were assigned to the younger age group for the 1967-70 period.

Irregularities in the estimates for the various periods of arrival suggest further difficulties with the estimation procedure. The estimated undercounts "in 1970" for foreign-born Cubans who arrived prior to 1960 represent bicensal relative net undercounts. Thus, the apparent slight overcount for females of about 1 percent for this group implies an improvement in coverage for the group. For the group arriving in the 1960-65 period, the estimates are clearly implausible. The 10-percent overcount for females coupled with a 2-percent undercount for males for this period suggests some possible misallocation by sex and/or by time period in estimating the expected population. This interpretation is presumably supported by the large apparent undercount for each sex in the 1965-70 period, in spite of the arbitrary assignment of most Cubans residing in Puerto Rico to this period and their exclusion from the expected number. Some of the immigrants assigned to the 1965-70 period should presumably have been assigned to the earlier period.

The estimates of census coverage in 1970 for foreign-born Cubans shown in table 15 can be improved by correcting the 1960 census counts for coverage errors. However, since coverage estimates for this population are not available for 1960, it was arbitrarily assumed for illustrative purposes that the 1960 undercount rates (for age and sex categories) for the Black-and-other-races population also applied to the foreign-born Cuban population. With this assumption (implying an undercount of 9.7 percent in 1960), the resulting estimated overall undercount rate of foreign-born Cubans in 1970 would be about 4.6 percent, or 6.7 percent for males and 2.7 percent for females. Unlike the population of Puerto Rican birth or parentage, the estimated undercount for foreign-born Cubans in 1970 is not very sensitive to the assumed coverage rates in 1960. If the 1960 rates are assumed to be twice those for Blacks-and-other races (or 20.3 percent), the estimated overall undercount rate in 1970 for foreign-born Cubans would increase to only 6.8 percent (figure 4). This lack of sensitivity to the assumed 1960 coverage rates can be attributed to the fact that about five-sixths of the foreign-born Cuban population in 1970 entered the United States after 1960.

It is useful to stress a few points about these calculations. The reliability of coverage estimates depends strongly on the reliability of the data and assumptions on which the estimates are based. Since the estimates of coverage for foreign-born Cubans in the 1970 census derived by analytic techniques are based largely on rather inadequate data and untested assumptions, they are only rough estimates at best. Consequently, the estimates, particularly those for age-sex categories, should be used with extreme caution. This exercise may be viewed as illustrative of some of the problems in estimating census coverage from deficient and limited data.

[58] More detailed calculations show that most of the shortage in the expected population is in the age group 5-9 and that there is a slight excess in ages 0-4.

Intercensal Cohort Analysis: Native Population of Mexican Parentage

Data on the native population of Mexican parentage from both the 1960 and 1970 censuses can be employed to prepare estimates of bicensal relative coverage error for this group by means of intercensal cohort analysis. If the 1960 census counts are not corrected for underenumeration and if the survivors in 1970 of the 1960 population account for all of the expected 1970 population (that is, if the expected population relates only to ages 10 and over in 1970), the differences between the expected and census populations in 1970 represent the differences between the errors in the 1970 counts and the errors in the 1960 counts for the cohorts already born by 1960 (that is, the bicensal relative error) rather than net undercounts in 1970 as such.

Problems in dealing with immigration, such as those encountered in the analysis of data on Cubans and Puerto Ricans, are avoided, or at least reduced, by restricting the analysis to natives. Since the net movement of natives into or out of the country between 1960 and 1970 was probably quite small, the net migration component was ignored in the present calculations. Mainly for this reason, intercensal cohort analysis is much simpler for natives of Mexican parentage than for the foreign-born Cuban population or the population of Puerto Rican birth or parentage.

The 1960 census counts of natives of Mexican parentage were published only in broad age groups for each sex. They were first subdivided into 5-year age groups on the basis of the age distribution of the White Spanish-surname population of Mexican parentage residing in the five Southwestern States in 1960. The expected population in 1970 was then derived by applying 10-year (time interval) survival rates from the life table for the Spanish-surname population of Texas in 1970 to the 1960 census population. Finally, the expected population in 1970 was compared with the 1970 census counts. As previously stated, no allowance was made for civilian immigration or emigration during the decade. Both the 1960 and 1970 populations were adjusted to include male members of the Armed Forces overseas.

One possible inference to be drawn from the intercensal cohort analysis for the native population of Mexican parentage is that coverage of this group improved between the two censuses by about 6 percent for each sex (table 16). Much of this apparent improvement in coverage occurred in only a few cohorts. Coverage of the cohorts aged 10–19 and over 70 in 1970 (aged 0–9 and over 60 in 1960) appears to have improved by over 20 percent between 1960 and 1970. These apparent coverage gains are so large that they cast doubt on the validity of the basic data. One likely source of error in the figures is misreporting of nativity by respondents for themselves and members of their households; e.g., in 1970, some persons born in Mexico who entered the country illegally may have reported a place of birth in the United States in order to legitimate their illegal presence in the United States or to bolster claims to American citizenship. Other potential sources of error are possible bias on the part of the Census Bureau in allocating nativity for persons who did not report place of birth and the omission of an allowance for net migration during the 1960-70 decade.

Table 16. Estimates of Bicensal Relative Error for the Native Population of Mexican Parentage, by Sex: 1960-1970

(Population in thousands)

Age (years)		Male				Female			
		Census[1], 1960 (1)	Expected population,[2] 1970 (2)	Census[1], 1960 (3)	Percent relative error[3] [(3)-(2)]÷(2)= (4)	Census[1], 1960 (5)	Expected population,[2] 1970 (6)	Census[1], 1970 (7)	Percent relative error[3] [(7)-(6)]÷(6)= (8)
In 1960	In 1970								
All ages	10 and over	582.9	558.0	593.1	+6.3	581.3	564.9	601.9	+6.5
0 to 4	10 to 14	72.4	71.8	88.6	+23.5	69.5	69.0	86.1	+24.8
5 to 9	15 to 19	64.8	64.4	76.2	+18.3	63.8	63.5	75.2	+18.4
10 to 14	20 to 24	64.3	63.6	62.3	-2.0	62.4	62.0	62.0	-0.1
15 to 19	25 to 29	58.1	57.2	54.1	-5.3	58.9	58.4	55.5	-5.0
20 to 24	30 to 34	53.4	52.2	57.9	+11.0	53.5	52.9	58.6	+10.7
25 to 29	35 to 39	61.3	59.5	62.8	+5.5	61.7	60.9	65.4	+7.5
30 to 34	40 to 44	68.4	66.0	66.6	+0.8	70.4	69.0	68.2	-1.1
35 to 39	45 to 49	51.4	49.2	50.2	+1.9	54.2	52.7	52.0	-1.4
40 to 44	50 to 54	31.7	29.8	28.5	-4.4	32.5	31.2	30.9	-1.0
45 to 49	55 to 59	20.9	18.9	18.5	-1.9	20.1	18.8	19.1	+1.4
50 to 54	60 to 64	13.2	11.2	11.3	+0.6	12.6	11.4	11.3	-0.9
55 to 59	65 to 69	9.0	7.1	7.6	+6.5	8.7	7.5	7.3	-1.8
60 to 64	70 to 74	5.1	3.6	4.2	+17.4	4.9	3.8	4.7	+25.1
65 and over	75 and over	8.8	3.6	4.4	+23.1	8.3	3.9	5.6	+46.4

[1]Adjusted to include Armed Forces overseas.
[2]Derived by applying life table survival rates for 1970 Spanish-surname population of Texas to 1960 census counts.
[3]Percentages computed from unrounded figures. Base of percent is expected population. A plus sign denotes a relative net overcount in the 1970 census as compared with the 1960 census; a minus sign denotes a relative net undercount.

Source of census figures: U.S. Bureau of the Census, Census of Population: 1970, Subject Reports, PC(2)-1A, National Origin and Language, 1973, table 10, p. 70. U.S. Bureau of the Census, Census of Population: 1960, Subject Reports, PC(2)-1A, Nativity and Parentage, 1963, table 9, p. 32.

The patterns of the relative coverage errors for the cohorts aged 20 to 69 in 1970 (10 to 59 in 1960) are indicative of certain types of coverage problems in the censuses even though the magnitude of the absolute coverage errors in 1970 cannot be ascertained. The relative overcount of males at ages 65–69 in 1970 suggests possible age missstatement in 1970 on the part of individuals who had not actually turned 65. The relative undercounts at ages 20 to 29 for both sexes, coupled with large relative overcounts at ages 30 to 39, suggest that the cohorts aged 20 to 29 in 1970 (as well as the cohorts aged 20 to 29 in 1960) had worse coverage than the adjacent age groups. For the remaining cohorts (males aged 40 to 64 in 1970, except 50 to 54; females aged 40 to 69) the relative coverage errors are quite small, possibly smaller than the sampling error of the census data; this pattern indicates stability in census coverage over time.

As previously noted, to derive absolute measures of coverage error for the native population of Mexican parentage by intercensal cohort analysis, it is necessary to correct the 1960 census counts for coverage errors. No direct measures of census coverage in 1960 are available for this population so that it is not possible to produce definitive coverage estimates for it in 1970. However, illustrative estimates of coverage error in 1970 can be derived by assuming the level of coverage errors for 1960. If it is assumed, for example, that coverage rates by age and sex for the native population of Mexican parentage in 1960 were the same as the rates for Black-and-other-races population in 1960 (implying an overall undercount rate of 9.2 percent), then the overall net underenumeration for natives of Mexican parentage in 1970 (aged 10 years and over) would be 3.4 percent.

The sensitivity of this result to varying assumptions for 1960 can be tested by assuming that various multiples of the Black-and-other-races coverage rates in 1960 apply to the 1960 native population of Mexican parentage. According to figure 4, the 1970 undercount rate for natives of Mexican parentage would exceed that for Whites (1.9 percent) if the 1960 undercount rates were greater than 0.84 times those of the Black-and-other-races population in 1960, or 7.7 percent overall. On the other hand, if the 1960 undercount rates for natives of Mexican parentage were less than 1.41 times those of the Black-and-other-races population, or 12.9 percent overall, then the undercount rate for natives of Mexican parentage in 1970 would be less than 7.7 percent, the undercount rate for Blacks in 1970. Thus, for a fairly wide range of assumed coverage rates in 1960, the undercount for natives of Mexican parentage falls between that for Whites and that for Blacks.

These illustrative estimates are subject to the several types of errors discussed earlier as well as the possibility that the pattern and level of coverage rates in 1960 for natives of Mexican parentage differed considerably from those of the Black-and-other-races population. Perhaps the only definitive statement that can be made on the basis of this analysis is that the coverage of natives of Mexican parentage probably improved by about 6 percent between 1960 and 1970.

Stable Population Models

Stable population models have been widely used in the evaluation of census data. Stable population models are theoretical, unchanging age distributions which result from and correspond to specific rates of population growth and specific schedules of fertility and mortality. Given the rate of growth of a population and its mortality level, for example, it is possible to specify its age distribution under certain conditions. The stable population model assumes unchanging fertility and mortality and the absence of net immigration over a long period. Under these circumstances the stable age distribution corresponding to the observed growth rate and the observed fertility and mortality rates may be taken to represent the true age distribution for purposes of evaluating the enumerated age distribution. The Hispanic population does not conform to this model, however, viewed from any aspect, now or in the past. Another type of population model, the quasi-stable model, assumes constant fertility, slowly declining mortality, and the absence of net immigration; this model also does not fit the case of the Hispanic population.

These model age distributions essentially correct for errors of age reporting; they correct for coverage errors to only a minimal extent. This inference follows from the fact that stable population analysis does not in itself provide for an alternative total for a population, only for an alternative age distribution. Coverage estimates made for specific age-sex groups, using these or similar models, are likely to be very much in error.

Comparison With Aggregate Administrative Record Data

Another possibility for obtaining coverage estimates for the Hispanic population involves comparing counts from a set or sets of administrative records according to one of the identifiers with counts from the census. The records must be complete or must be adjusted for incomplete coverage, and the Hispanic identifier must be the same for the administrative records as for the census. A further requirement is that the records be national in scope or, if restricted to a particular geographic area, contain valid addresses as of the census day. We have not been able to identify any such sets of administrative records for use in estimating the coverage of the Hispanic population in the 1970 census.

The Hispanic identifier which could prove to be most useful for exploiting administrative records is Spanish surname. It is a quite objective criterion, compared with Spanish origin and similar subjective criteria. Furthermore, the use of Spanish surname does not require that persons identify themselves as Hispanic upon entry into the record system; that is, it is not necessary for the agency which created the record system to have included identification of individuals as Hispanic in its original objectives. The major drawback of the Spanish-surname identifier is that the list of names, as now constituted, has not proven successful in identifying the Hispanic population outside the five Southwestern States.[59]

[59] U.S. Bureau of the Census, Technical Paper No. 38, op. cit.

Research is now in progress which could result in the development of a list of Spanish surnames for use in identifying Hispanic persons in administrative record files and in the census for the entire country. (See section, "Plans for Evaluation Studies in 1980".)

The Social Security Administration/Internal Revenue Service files could prove to be quite useful for estimating the coverage of the Hispanic population even though, at present, neither the application form for a Social Security number nor the other record forms of the SSA or the IRS contain explicit information identifying Hispanic persons. One approach to the problem would involve matching a sample of persons with Spanish surnames in the SSA/IRS files to the census for 1970. However, this procedure is costly and complex and is not now practicable for 1970.

CONCLUSION

Development of definitive estimates of coverage of the Hispanic population in the 1970 census has not proved to be feasible at this time, on the basis of any of the techniques discussed in the preceding sections. However, some very general indications of coverage levels and patterns have been identified. The general measures of errors in the age and sex data of the Hispanic population suggest that the coverage level of the Hispanic population in 1970 falls between that of the White and Black populations.[60] Furthermore, deficiencies in certain age and sex groups are apparent. For example, Hispanic males are not enumerated as completely as females; undercoverage of young adults, particularly males, is relatively high; and there are notable reporting problems among the elderly. Estimates of coverage for three national-origin subgroups of the Hispanic population—foreign-born Cubans, Puerto Ricans, and the native population of Mexican parentage—strongly suggest that coverage for these groups taken separately and in combination (constituting less than two-fifths of the reported population of Spanish origin in 1970) falls between that of the White and Black populations. However, in view of the lack of appropriate data, specific estimates of coverage for the entire Hispanic population in 1970 would have to be based on a large number of invented assumptions of questionable validity. Such estimates could possibly be worthless and even very misleading.

[60] The only direct evidence currently available regarding the coverage of the Hispanic population comes from coverage studies conducted in conjunction with two pretests for the 1980 census. The evidence from these studies is not inconsistent with the hypothesis

[60]—Continued

that undercoverage of the Hispanic population is intermediate between that of Whites and Blacks. The available results relate to coverage of housing units, however, so that the figures may not reflect the relative levels of coverage of population closely. In the coverage studies conducted following the censuses of Travis County, Texas and Camden, New Jersey, each taken in 1976, enumeration districts (ED's) were designated as either Spanish (i.e., over 30 percent of the residents were of Spanish origin) or non-Spanish and a sample of blocks was chosen from each group. In Travis County, there was no significant difference (at a 95-percent confidence level) between miss rates for housing units in Spanish areas (1.6 percent) and non-Spanish areas (0.6 percent). In Camden, the miss rate for housing units in Spanish areas (1.0 percent) was significantly lower than in non-Spanish areas (1.7 percent). (Camden's non-Hispanic population in 1976 was 48.3 percent Black.) These results are from T.W. Harahush and I. Fernández, "The Coverage of Housing Units—Results from Two Census Pretests," **Proceedings of the Social Statistics Section, 1978,** American Statistical Association.

Prospects for Measuring the Coverage of the Hispanic Population

PROSPECTS FOR DEVELOPMENT OF DATA

Our attempt to apply techniques of demographic analysis to estimate the coverage of the Hispanic population in the 1970 census has disclosed two areas in need of improvement before this approach can be expected to yield accurate estimates of census coverage for the Hispanic population or its subgroups. These are (1) a precise and consistent definition of the Hispanic population and (2) data from non-census sources, such as vital statistics, immigration data, and administrative record data, which are consistent with the census definition. These problems may continue to plague efforts to evaluate coverage of the Hispanic population in the 1980 census.

For the 1980 census, the Bureau of the Census plans to use a completely self-designating or subjective method of defining the Hispanic population. A person will be classified as Hispanic if he or she chooses a Hispanic category in response to the question on Hispanic origin or descent which is to appear on the 100-percent census questionnaire. Inasmuch as data on Hispanic origin or descent are also available from the 1970 census, the Census Bureau will have data on the Hispanic population based on an essentially common general definition for the last two censuses.[61] Even so, derivation of coverage estimates for the Hispanic population by the demographic method will continue to be difficult because of possible differences between these and other (non-census) figures on the Hispanic population resulting from shifts in the identification of individuals as Hispanic or non-Hispanic.

A division of the total "error" into the coverage and response components may be attempted through match studies in 1980. We are concerned, however, that the attempt may not result in reliable estimates of the components. Match studies may also be used to provide estimates of gross and net response error for the Hispanic population in 1980. They are likely to show that response variability accounts for a substantial portion of the measured error. Thus, although self-designated Spanish origin or descent may be an appropriate choice as the primary identifier of the Hispanic population, it will present difficulties in evaluating coverage in 1980.

Demographic data independent of the census corresponding to the concept of Hispanic origin or descent will be needed in order to evaluate the census figures on the Hispanic population in 1980 by the demographic method. Vital statistics and immigration data following this concept over a long period will be required. One method of obtaining the requisite vital statistics is simply to ask a question about Spanish origin or descent on the birth and death certificates. The National Center for Health Statistics has proposed to the States that they modify their birth and death certificates to secure information on Spanish origin. So far, 18 States, generally those with the largest numbers of persons of Spanish origin, have agreed to secure such information. The first data to be tabulated from these new certificates will relate to 1978 and may become available in 1980. Data are needed for the other States and for a long period a years, however, if they are to be useful for coverage evaluation. Although such data would not be strictly comparable to census data because the concept is subjective and the (same or different) respondent would not necessarily have provided the same information for the census and the birth or death certificates, they may be sufficiently consistent to serve as a rough basis for census evaluation in the coming years for at least some portion of the Hispanic population.

Another possibility for obtaining vital statistics on the Hispanic population—one which does not involve any modification of the present birth and death certificates—is the coding of birth and death certificates according to Spanish surnames. Coding of Spanish surnames has the distinct advantage that it is relatively objective and can be applied retrospectively, making it possible to produce a series of consistent data for a number of years. These data could then be used to evaluate the census coverage of the Spanish-surname population in the younger ages in 1980.[62] This procedure would have a number of drawbacks; these arise primarily from the lack of correspondence between Hispanic origin and Spanish surname as identifiers, as noted earlier. Moreover, estimates of the completeness of birth registration for this group are lacking.

Any application of the component method to the estimation of the coverage of the Spanish-surname population would require, in addition to the appropriate vital statistics, national data on net immigration or regional data

[61] The question on Spanish origin has been revised for the 1980 census, from the 1970 census question, "Is this person's origin or descent . . .?" to read "Is this person of Spanish/Hispanic origin or descent?" This new format is expected to reduce errors caused by misunderstanding of the question and to produce more consistent data.

[62] At this time (April 1979), it appears unlikely that Spanish surname will be coded outside the five Southwestern States in the 1980 census.

on net in-migration (e.g., State-of-birth data classified by age) for the five Southwestern States for the Spanish-surname population. Calculating coverage estimates for the entire United States with national data would alleviate this problem, but the problem of applying the present Spanish-surname list to the census data and the vital statistics for areas outside the Southwest would remain.

Coding of Spanish surnames could also be done (perhaps on a sample basis) for some sets of administrative records, such as Social Security summary earnings files, Medicare enrollments, or IRS tax returns for 1980. It would then be possible to compare the aggregate census figures for persons with Spanish surnames resident in the five Southwestern States or in the United States with similar figures from the Medicare files or the Social Security files. Aggregate comparison with tabulations from the Social Security summary earnings files has severe limitations, however, resulting principally from omissions from and duplications in the files. The coded records could be used more effectively in a matching study, which does not depend on the completeness of the file. For example, a sample of persons with Spanish surnames in the SSA/IRS files could be matched to the 1980 census records to evaluate the coverage of selected age groups in the Hispanic population.

In any application involving the use of a list of Spanish surnames to identify the Hispanic population, the many problems associated with this identifier, previously discussed, must be overcome. Research at the Census Bureau is now under way in an attempt to resolve some of these problems. Refining the Census Bureau's list so that it could serve as a standard list for use by the National Center for Health Statistics, the Social Security Administration, and the Immigration and Naturalization Service, as well as by the Census Bureau, is one goal of the research. The refined list may aid in the tabulation of the births and deaths of the Hispanic population and in the calculation of the corresponding birth and death rates, as well as in directly measuring coverage by such techniques as a case-by-case match with Social Security files. The research is also proceeding in the direction of adapting the list, or devising a second one, for use in identifying the Spanish-surname population outside the southwest.[63]

Representatives of the Census Bureau and the Social Security Administration have been collaboratively exploring the issues involved in securing Social Security data classified by Spanish ancestry. The Census Bureau has received a tabulation of a 1-percent sample of persons included in the Social Security file and has attempted to identify those persons with Spanish surnames as a test of the feasibility of making this classification in 1980. The Census Bureau has also indicated to the Social Security Administration its interest in having the application form for a Social Security number revised to include identification of Spanish origin or descent.

The number of illegal aliens in the United States is a subject which has received a great deal of attention in bureaucratic, political, professional, and popular writings and discussions. The range of the available estimates for recent years is quite large, from 1 to 12 million, and the methods used to derive the estimates are quite varied, ranging from conjecture based on enforcement data[64] to statistical inference based on comparison of time series of aggregate administrative data,[65] triple-system match studies,[66] and demographic analyses of survey data at various dates.[67] A direct survey has also been undertaken.[68] The analytical research to date does suggest that the number of illegal aliens in the country and the *net* flow are far less than the highest conjectural estimates.[69]

In a further attempt to apply demographic analysis to the estimation of the number of illegal aliens in the country, the Census Bureau has undertaken an analysis of trends in age-sex-cause-specific death rates for the group of States in which illegal aliens are believed to be numerous. This analysis also suggests that the number of illegal aliens in the country is probably not as large as the highest conjectural estimates and that large increases in the *net* flow of illegal aliens have not occurred since 1970.[70]

To date, this and other research projects have not yielded any definitive or preferred estimate of the net flow or current population of illegal aliens. The question as to how many illegal aliens are in the country remains open.

---

[63] The basic data for this research is a 20-percent sample of all Social Security records. From this file, it is possible to tabulate the frequency of occurence of every surname (Spanish and non-Spanish) in each State. Spanish surnames can then be identified on the basis of geographic (distributional) criteria as well as the usual linguistic and genealogical criteria. For a more detailed discussion of the research on Spanish surnames, see D.L. Word, J.S. Passel, B.D. Causey, and E.W. Fernandez, "Determining a List of Spanish Surnames by Analysis of Geographical Distributions," unpublished paper presented at the annual meeting of the Southern Regional Demographic Group, San Antonio, Texas, October 1978.

[64] Lesko Associates, Final Report: **Basic Data and Guidance Required to Implement a Major Illegal Alien Study During Fiscal Year 1976,** prepared for the Office of Planning and Evaluation, U.S. Immigration and Naturalization Service. Washington, D.C., October 1975.

[65] Alexander Korns, "Coverage Issues Raised by Comparisons Between CPS and Establishment Employment," **Proceedings of the Social Statistics Section, 1977: Part I,** American Statistical Association, 1978, pp. 60-69.

[66] Clarise Lancaster, and Frederick J. Scheuren, "Counting the Uncountable Illegals: Some Initial Statistical Speculations Employing Capture-Recapture Techniques," **Proceedings of the Social Statistics Section, 1977: Part I,** American Statistical Association, 1978, pp. 530-535.

[67] David M. Heer, "What is the Annual Net Flow of Undocumented Mexican Immigrants to the U.S.?", paper presented at the annual meeting of the Population Association of America, Atlanta, Georgia, April 13-15, 1978.

[68] J.A. Reyes Associates, **The Survey Design for a Residential Survey of Illegal Aliens,** report submitted to the Immigration and Naturalization Service, September 5, 1977. See also Statement of Robert Warren, pp. 704-706, in **Immigration to the United States,** Hearings before the Select Committee on Population, U.S. House of Representatives, Ninety-fifth Congress, Second Session, April 4-7, 1978, No. 5,U.S. Government Printing Office, Washington, D.C. 1978.

[69] Select Committee on Population, U.S. House of Representatives, Ninety-fifth Congress, Second Session, **Legal and Illegal Immigration to the United States,** Serial C, December 1978. See also **Immigration to the United States, op. cit.** and Charles B. Keely, "Counting the Uncountable: Estimates of Undocumented Aliens in the United States." **Population and Development Review,** Vol. 3, No. 4, December 1977, pp. 473-482.

[70] J. Gregory Robinson, "Estimating the Approximate Size of the Illegal Alien Population in the United States by the Comparative Trend Analysis of Age-Specific Death Rates", paper presented at the annual meeting of the Population Association of America, Philadelphia, Pennsylvania, April 26-28, 1979.

PLANS FOR EVALUATION STUDIES IN 1980

Plans for the evaluation of the coverage of the population in the 1980 census are still being formulated. These plans envisage an effort to evaluate the coverage of the Hispanic population of the United States, States, and possibly constituent geographic areas. It is planned to apply the various techniques of demographic analysis considered here for 1970 as well as extensions and adaptations of them that the data permit. The scope of this approach will necessarily remain limited, however. In spite of the steps being taken to expand the range of Hispanic data, particularly vital statistics, these will be of little utility for 1980 since generally the States are only beginning to collect data on births and deaths of Hispanic persons.

At present, principal consideration is being given to the conduct of a post-enumeration survey (PES), the results of which would be matched to the census records on a case-by-case basis. Individuals would be identified as of Spanish origin or descent in the post-enumeration survey and this determination would establish their identification as Hispanic for the purposes of the study. It is quite possible that the census-PES match study will provide satisfactory estimates of coverage for the Hispanic population at most for the United States as a whole and that other methods will be required to produce estimates for the major political subdivisions of the United States.

Information from the match study on the variation in the coverage rates of the Hispanic population according to socioeconomic characteristics for the United States may be employed in a synthetic or regression design, in combination with census data on the geographic and socioeconomic distribution of the Hispanic population, to prepare estimates of coverage for the Hispanic population of the major political subdivisions of the United States. Coverage rates according to socioeconomic characteristics for the general U.S. population may permit the calculation of coverage rates for the Spanish population in the United States as a whole in the event that satisfactory Hispanic coverage rates for the United States are not available from the match study.

A further match of the post-enumeration survey with the Social Security/Medicare/IRS files on the basis of Spanish surname is being considered. The results of such a study could be used to adjust the original census-PES estimate of coverage for understatement, that is, to allow for groups which tend to be omitted from censuses and surveys. Presumably the administrative files include a representation of persons living in units that were omitted from the census or improperly reported as vacant and persons whom householders carelessly or deliberately excluded from census reports. Hopefully, this device will serve to encompass the measurement of illegal aliens of Hispanic origin, even though this group cannot be identified separately.

Given the uncertainties of the triple-system matching scheme, the response variability in the definition of the Hispanic population, and the deficiencies in the various non-census data sources, it is not clear at this time whether reliable estimates of coverage of the Hispanic population can be developed nationally or at any subordinate geographic level for 1980. A major effort will be made to do so, however.

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Official Business
Penalty for Private Use, $300

FIRST-CLASS MAIL
POSTAGE & FEES PAID
CENSUS
PERMIT No. G–58