LINKING IRS AND CENSUS DATA TO PRODUCE INTERCENSAL INCOME ESTIMATES: AN EXPLORATION

Evelyn S. Mann and Joseph J. Salvo, New York City Department of City Planning

INTRODUCTION

At the present time, there are few, if any, intercensal surveys from which county or sub-county income estimates by household and family characteristics could be derived. The sample sizes of most intercensal surveys are usually too small to allow for reliable, detailed esti-mates at local geographic levels. Cancellation of plans to conduct a mid-decade census and ever-increasing financial austerity point in the direction of fewer alternative data sources.

One course of action is to look towards ad-ministrative data in lieu of sample surveys. Recently, some attention has been devoted to the possible use of Internal Revenue Service (IRS) income tax data as a basis for local level es-timates of income by household and family char-acteristics. The purpose of this paper is to describe a methodology currently being developed to create such estimates on a county and sub-county level in New York City. The goal of this research is to examine relationships between income data by household or family type, size and composition from the 1980 Census and con-ceptually comparable data from 1979 Federal income tax returns, in order to produce local census-type income estimates for 1984.

It should be noted at the outset that no amount of data manipulation can produce absolute conceptual comparability between Census and IRS data. This research does not seek to "match" IRS with Census data, but instead, attempts to superimpose the conceptual framework of census on IRS data, so that relationships may be ex-amined between approximately parallel concepts. If the ratios of census to IRS income variables prove to be consistent from area to area, then estimates of income by household or family char-acteristics could be derived between decades, using IRS data as a basis for such estimates. Of major concern is how the ratios may interact with pertinent variables, such as socioeconomic status or size of area. The presence of con-sistent or systematic variation in the ratios, as opposed to random variation, will ultimately determine the utility of this technique. In either case, however, this research will serve to show whether or not IRS data display poten-tial as a basis for income estimation.

DATA SOURCES

This research utilizes data for about three million Federal income tax returns for the cal-endar year 1979. The specific variables made available by the Internal Revenue Service are displayed in Appendix Table A. These are limited to filing status of returns, number and type of exemptions taken, and income by type. A recode operation will create the variables necessary for comparison with detailed tabulations from Summary Tape File 4 (STF-4).

Implementation of the recode operation will be performed under the strict supervision of the Internal Revenue Service at a state governmental site. At no time will researchers have access to individual tax returns. Data will be first recoded, then geocoded and aggregated for each of New York City's 59 Community Districts. These are the statistical areas used in the Neighborhood Statistics Program, with an average population of about 120,000 persons, the size of many central cities of metropolitan areas throughout the country.

Presently, specifications for an elaborate suppression routine are being designed and tested. Disclosure procedures mandated by the Internal Revenue Service require a suppression routine which prohibits any derivation of any cell with less than 10 cases or based on fewer than 10 cases.

DRAWING ANALYTICAL PARALLELS

The principal difficulty in any analytical comparison of census and IRS data is that units of analysis are inherently different. A Census questionnaire represents a household with either a family or a nonfamily group within or it enu-merates an individual in a group quarters setting. An IRS return, on the other hand, may or may not represent a whole household or family, as in the case of persons living with families or nonrelatives filing single returns. Despite this, it is believed that conceptual parallels may be drawn which permit data comparisons.

Basic parallels between household and family type, size and composition on census and IRS records are shown in Tables 1 and 2, reflecting some of the basic recodes performed on the data. In Table 1, at-home exemptions are equated with family size. The four types of exemptions listed in Table 1 which comprise the number of at-home exemptions represent a list of relatives which include the householder, spouse, children and parents of the householder. Since there is no definitive way of identifying whether "other" dependents are "at-home exemptions" or whether they are even relatives, for that matter, no attempt was made to expand the definition of family or measure household size in the IRS scheme. Therefore, this analysis excludes those "other" relatives more typical of extended families, such as uncles and aunts.

One of the more precise parallels which may be drawn is between children in families in the Census tabulation and children-at-home exemptions on IRS returns. The only real difficulty with such a comparison is that the census income tables in STF-4 refer to families with "own children," while information on children from IRS returns is really more akin to the "related children" concept. However, of the more than 1.7 million children under the age of 18 in New York City, only about seven percent are

TABLE 1.--INDICATORS OF FAMILY/HOUSEHOLD SIZE AND COMPOSITION: PARALLELS BETWEEN CENSUS AND IRS[1]

CENSUS (STF-4)	IRS

Family Size
  2 Persons
  3 Persons
  .....
  6 or More Persons
	Sum of At-home Exemptions:
	(a) Primary Taxpayer (excluding age/blindness)
	(b) Secondary Taxpayer (excluding age/blindness)
	(c) Taxpayer's Children-at-home
	(d) Taxpayer's parents-at-home

Persons in Families	Number of At-home Exemptions
Families with Own Children	Presence of Children-at-home Exemptions
Number of Own Children in Families	Number of Children-at-home Exemptions
Households/Families with Heads 65 and Over	Primary Taxpayer Exemptions More Than 1

[1]IRS family and household composition data are based on calendar year 1979; comparable Census data are as of April, 1980.

NEW YORK CITY DEPARTMENT OF CITY PLANNING POPULATION DIVISION,   10/83

classified as related children who are not own children of the householder. Since a taxpayer may claim more than just his own children (e.g., children living in the family whose relationship to the householder is that of a grandchild or nephew), it is expected that IRS returns should overestimate total children relative to the census count of "own children."

In this study all exemptions taken for age or blindness are excluded from computations and analyses. The only exception to this is shown in Table 1, where returns are checked to see if there is more than one exemption in the primary taxpayer exemption field. Since, nationally, only 1.5 percent of all exemptions for age and blindness were actually taken for blindness (Appendix Table B), more than one exemption in the primary taxpayer field is taken to mean that the householder is age 65 or over. This permits one to flag returns and construct tables just for households with aged heads. Planned table comparisons for households with householders 65 and over will parallel many of those presented for all families.

HOUSEHOLD AND FAMILY TYPOLOGY

The family typology recode, which serves as the basis for data analysis, is presented in Table 2. The most straightforward comparison is that for married-couple families. It is recognized that the inclusion of married persons filing separate returns will probably cause some overestimation of married-couple families in the IRS data base. The fact that under four percent of all 1979 national returns filed by married couples fell within this category (Appendix Table B) should serve to attenuate this effect.

The "other family" category represents an attempt to examine nontraditional family situations. In New York City, this is particularly important, given that 32 percent of all families are nontraditional, that is, having only a male

or female householder without a spouse present. Unfortunately, the sex of the householder cannot be determined from the IRS data file (Appendix Table A), nor can any attempt be made to distinguish between different types of nontraditional families, such as two or more related adults living together, e.g., mother-son, sister-sister, etc.

The most problematic of all household and family types is the "single, one at-home exemption" group. The number of persons in this category probably represents an overestimate of both the number of one-person households and the number of unrelated individuals. Indeed, Irwin and Herriot (1982), comparing national IRS with Current Population Survey (CPS) data, found that the number of returns filed in 1980 reporting only one exemption was more than double the number of one-person households reported on the March 1980 CPS. While many persons filing "single, one at-home exemption" are truly living alone, other, such as young students living with families or nonrelatives, are not. In the same vein, not all persons who are filing as "single, one at-home exemption" live with nonrelatives. The IRS estimate includes singles who cannot be considered unrelated individuals.

COMPARISON TABLES: STF-4 AND IRS

Several key comparisons being made between data from STF-4 and IRS records are shown in Tables 3 through 8. By combining the IRS family typology recode (Table 2) with information on the number and type of exemptions (Table 1), cross-tabulations will be constructed which are roughly comparable to tables from STF-4. In Table 3, for example, the number of married-couple families and "other" families by presence of own children will be compared to the frequency of family returns by family type and presence of children at-home exemptions from the IRS database. This comparison is particularly

TABLE 2.- - FAMILY/HOUSEHOLD TYPOLOGY: PARALLELS BETWEEN CENSUS AND IRS[1]

CENSUS (STF-4)	IRS FILING STATUS
Married:	**Married:**
Married-couple Families	Married, Filing Jointly
	Married, Filing Separately
Other Family:	**Other Family:**
Females, No Spouse, with Children	Unmarried Head of Household
Males, No Spouse, with Children	Surviving Widow(er) with Dependent Children
Other Family, No Children	Single, Total at-home Exemptions[2] More than 1
Single:	**Single:**
Single Person Households/ Unrelated Individuals	Single, Total at-home Exemptions[2] = 1

[1]IRS family and household composition data are based on calendar year 1979; comparable Census data are as of April, 1980.

[2]For a listing of exemptions included in the "At-home" grouping, see Table 1.

NEW YORK CITY DEPARTMENT OF CITY PLANNING   POPULATION DIVISICN,   10/83

meaningful given the fact that nontraditional families, those in the "other family" category, with children are a major focus of policy issues and social programs.

By examining the frequency of at-home exemptions, an indication of the population living in different household and family situations will also be obtained (Table 4). Estimates in IRS terms of the number of persons living in married-couple and in "other" families will be obtained by examining the aggregate number of at-home exemptions in each filing status category and will be compared to the number of persons in households, by household type. These will be compared to census Table PB19. Moreover, the number of children living in each family group will be obtained by counting the frequency of at-home exemptions tabulated for children at-home and then compared with census counts of own children by type of family. As part of this comparison, returns classified as "single, one at-home exemption" will be compared

TABLE 3. - - FAMILY TYPE BY PRESENCE OF OWN CHILDREN: PARALLELS BETWEEN CENSUS AND IRS[1]

CENSUS (STF-4)	IRS
Table PB18 Family Type by Presence and Age of Own Children	Number of Returns by Filing Status Type and Presence of Children-at-home Exemptions
Universe: Families	Universe = Family Returns
Married-couple Family With Own Children Without Own Children	Married-couple Family Children-at-home Exemptions Present Children-at-home Exemptions Absent
Family with male householder, no wife present With Own Children Without Own Children	
Family with female householder, no husband present With Own Children Without Own Children	Other Family Children-at-home Exemptions Present Children-at-home Exemptions Absent

[1]IRS family and household composition data are based on calendar year 1979; comparable Census data are as of April, 1980.

NEW YORK CITY DEPARTMENT OF CITY PLANNING   POPULATION DIVISION,   10/83

TABLE 4--PERSONS BY FAMILY/HOUSEHOLD TYPE: PARALLELS BETWEEN CENSUS AND IRS[1]

CENSUS(STF-4)	IRS
Table PB19 Relationship and Household Type	Number of Exemptions by Filing Status and Exemption Type
Universe: Number of Persons in Households	Universe = Aggregate At-home Exemptions

Total in Households
Relatives (in families)
  In married-couple family
  In other family
	In Married-couple Families
	
	In Other Families

Nonfamily householders and nonrelatives (Unrelated individuals in households)	
Table PB7 Persons in Household	— Single
Universe: Households	
1 Person	

Table PB17 Family Type	
Universe: Own Children	
Married-couple family Family with male householder, no wife present	Children at-home in married-couple families
Family with female householder, no husband present	Children at-home in other families

[1] IRS family and household composition data are based on calendar year 1979; comparable Census data are as of April, 1980.

NEW YORK CITY DEPARTMENT OF CITY PLANNING   POPULATION DIVISION,   10/83

to both the number of one-person households and the number of unrelated individuals in households. In Table 5, the number of families by size of family will be compared with all family returns from the IRS records by number of at-home exemptions. Since family size by family type is not available on STF-4, any such detailed comparison with IRS data is precluded.

TABLE 5.--FAMILY SIZE: PARALLELS BETWEEN CENSUS AND IRS[1]

CENSUS (STF-4)	IRS
Table PB8 Persons in Family	Number of Returns by Family Size
Universe: Families	Universe = Family Returns
All Families	All Family Returns Number of At-home Exemptions
2 Person Families	Two
3 Person Families	Three
4 Person Families	Four
5 Person Families	Five
6 or More Person Families	Six
7 or More Person Families	

[1] IRS family and household composition data are based on calendar year 1979; comparable Census data are as of April, 1980.

NEW YORK CITY DEPARTMENT OF CITY PLANNING
POPULATION DIVISION,   10/83

It should be kept in mind that although Census and IRS income data are based on the calendar year 1979, the time frame surrounding household and family items is slightly different between data sets. Filing status and exemption type on IRS returns are for the 1979 calendar year, while the household and family items on census returns are as of the census enumeration date, April 1980, for most households. In all comparisons, IRS figures represent adjusted-gross income, whereas those on census returns theoretically represent total money income. This is not necessarily the case with IRS, where filing is not required below a certain level of income or where most income is from a tax-exempt source, such as social security. This should result in smaller estimates of households and families from IRS data in low income areas, as well as a deficit of households and families reporting very low income in all areas.

Table 6 focuses on the comparison of family income with the income of family returns by family type and presence of children-at-home exemptions from IRS records. This table may prove to be the most valuable, since income estimates are derived for single-parent families with children. In Table 7, the income distribution of "single, one at-home exemption" returns is compared with that of unrelated individuals. Unfortunately, a comparison table on the income of one-person households is not available from STF-4. This shortcoming points to an important problem on the Census side of this compar-

298

TABLE 6 --INCOME BY FAMILY TYPE AND PRESENCE OF OWN CHILDREN: PARALLELS BETWEEN CENSUS AND IRS

CENSUS (STF-4)[1]	IRS[2]
Table PB85 Family Type by Presence and Age of Own Children by Family Income in 1979	Income by Filing Status Type and Presence of Children-at-home Exemptions
Universe: Families	Universe = Family Returns

CENSUS (STF-4)[1]	IRS[2]

Married-couple family
  With own children
    Less than $ 2,500
    $ 2,500 to $ 4,999
    $ 5,000 to $ 7,499
    $ 7,500 to $ 9,999
    $10,000 to $12,499
    $12,500 to $14,999
    $15,000 to $17,499
    $17,500 to $19,999
    $20,000 to $22,499
    $22,500 to $24,999
    $25,000 to $27,499
    $27,500 to $29,999
    $30,000 to $34,999
    $35,000 to $39,999
    $40,000 to $49,999
    $50,000 to $74,999
    $75,000 or more
  Without own children
    (Repeat family income in 1979)
 	
Family with male householder,
  no wife present
  (Repeat presence of own children
  by family income in 1979)
 	
Family with female householder,
  no husband present
  (Repeat presence of own children
  by family income in 1979)
	Married-couple family
	  Children-at-home exemptions present
	    Less than $ 2,500
	    $ 2,500 to $ 4,999
	    $ 5,000 to $ 7,499
	    $ 7,500 to $ 9,999
	    $10,000 to $12,499
	    $12,500 to $14,999
	    $15,000 to $17,499
	    $17,500 to $19,999
	    $20,000 to $22,499
	    $22,500 to $24,999
	    $25,000 to $27,499
	    $27,500 to $29,999
	    $30,000 to $34,999
	    $35,000 to $39,999
	    $40,000 to $49,999
	    $50,000 to $74,999
	    $75,000 or more
	  Children-at-home exemptions absent
	    (Repeat income categories)
	
	Other Family
	  (Repeat children at-home exemptions
	   present/absent and income categories)

[1]Census data on family type/composition are as of April, 1980. Income is total income from all sources for the calendar year 1979.

[2]IRS data on family type/composition and income are based on calendar year 1979. Income equals Adjusted Gross Income (after "Adjustments to income").

NEW YORK CITY DEPARTMENT OF CITY PLANNING   POPULATION DIVISION   10/83

TABLE 7---INCOME OF UNRELATED INDIVIDUALS: PARALLELS BETWEEN CENSUS AND IRS

CENSUS (STF-4)[1]	IRS[2]
Table PB92 Sex by Age by Income in 1979	Income by Filing Status Type
Universe: Unrelated Individuals 15 Years and Over[3]	Universe = Single Returns

Total
  Less than $ 1,000
  $ 1,000 to $ 1,999
  $ 2,000 to $ 2,999
  $ 3,000 to $ 3,999
  $ 4,000 to $ 4,999
  $ 5,000 to $ 5,999
  $ 6,000 to $ 6,999
  $ 7,000 to $ 7,999
  $ 8,000 to $ 8,999
  $ 9,000 to $ 9,999
  $10,000 to $11,999
  $12,000 to $14,999
  $15,000 to $24,999
  $25,000 to $49,999
  $50,000 or more
	Total
	  Less than $ 1,000
	  $ 1,000 to $ 1,999
	  $ 2,000 to $ 2,999
	  $ 3,000 to $ 3,999
	  $ 4,000 to $ 4,999
	  $ 5,000 to $ 5,999
	  $ 6,000 to $ 6,999
	  $ 7,000 to $ 7,999
	  $ 8,000 to $ 8,999
	  $ 9,000 to $ 9,999
	  $10,000 to $11,999
	  $12,000 to $14,999
	  $15,000 to $24,999
	  $25,000 to $49,999
	  $50,000 or more

[1]Census data on family type/composition are as of April, 1980. Income is total income from all sources for the calendar year 1979.

[2]IRS data on family type/composition and income are based on calendar year 1979. Income equals Adjusted Gross Income (after "Adjustments to income").

[3]Including some not in households.

NEW YORK CITY DEPARTMENT OF CITY PLANNING   POPULATION DIVISION   10/83

TABLE 8.--FAMILY INCOME BY SIZE OF FAMILY: PARALLELS BETWEEN CENSUS AND IRS

CENSUS (STF-4)[1]	IRS[2]
Table PA23 Persons in Family by Family Income in 1979	Income by Filing Status Type and Number of At-home Exemptions
Universe: Families	Universe = Family Returns (2 or More At-home Exemptions)
All Families	All Family Returns
2 Person Families Less than $ 2,500 $ 2,500 to $ 4,999 $ 5,000 to $ 7,499 $ 7,500 to $ 9,999 $10,000 to $12,499 $12,500 to $14,999 $15,000 to $17,499 $17,500 to $19,999 $20,000 to $22,499 $22,500 to $24,999 $25,000 to $27,499 $30,000 to $34,999 $35,000 to $39,999 $40,000 to $49,999 $50,000 to $74,999 $75,000 or more	Two exemptions Less than $ 2,500 $ 2,500 to $ 4,999 $ 5,000 to $ 7,499 $ 7,500 to $ 9,999 $10,000 to $12,499 $12,500 to $14,999 $15,000 to $17,499 $17,500 to $19,999 $20,000 to $22,499 $22,500 to $24,999 $25,000 to $27,499 $30,000 to $34,999 $35,000 to $39,999 $40,000 to $49,999 $50,000 to $74,999 $75,000 or more
3 person families (Repeat family income in 1979)	Three exemptions (Repeat income categories)
4 person families (Repeat family income in 1979)	Four exemptions (Repeat income categories)
5 person families (Repeat family income in 1979)	Five exemptions (Repeat income categories)
6 or more person families (Repeat family income in 1979)	Six or more exemptions (Repeat income categories)
7 or more person families (Repeat family income in 1979)	

[1]Census data on family type/composition are as of April, 1980. Income is total income from all sources for the calendar year 1979.

[2]IRS data on family type/composition and income are based on calendar year 1979. Income equals Adjusted Gross Income (after "Adjustments to income").

NEW YORK CITY DEPARTMENT OF CITY PLANNING   POPULATION DIVISION,   10/83

ison--the absence of income information for nonfamily households. Although income tabulations are presented for unrelated individuals, this is not a household concept. It would be helpful if the census tabulated income information for unrelated individuals by variables indicative of living arrangements (e.g., size of household, those living in households and those in group quarters). Finally, Table 8 compares the income distribution of all family returns by family size.

In addition to the income categories displayed in each of the tables, the aggregate income reported from the two sources will also be compared. The aggregate income estimates will be used to calculate the ratio of Census to IRS income and to evaluate whether the technique developed here shows true promise.

NOTE: Appendix tables mentioned in the text were omitted from this version of the paper due to page constraints. The full paper appears in Statistics of Income and Related Administrative

Record Research: 1984, Internal Revenue Service, October 1984, pp. 131-138.

REFERENCES

Irwin, Richard and Roger Herriot, "An Initial Look at Preparing Estimates of Household Size for Local Areas from Federal Income Tax Returns," 1982 American Statistical Association Proceedings, Section on Survey Research Methods.

U.S. Department of Commerce, Bureau of the Census, Census of Population and Housing, 1980: Summary Tape File 4 Technical Documentation, prepared by the Data User Services Division, Bureau of Census, Washington, DC, 1983.

U.S. Department of the Treasury, Internal Revenue Service, Statistics of Income - 1979, Individual Income Tax Returns, U.S. Government Printing Office, 1982.