

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Income Size Distributions in the United States, Part I

Volume Author/Editor: Conference on Research in Income and Wealth

Volume Publisher: NBER

Volume ISBN: 0-870-14161-9

Volume URL: <http://www.nber.org/books/unkn43-2>

Publication Date: 1943

Chapter Title: Recommendations Designed To Reduce Heterogeneity Of Data

Chapter Author: Conference on Research in Income and Wealth

Chapter URL: <http://www.nber.org/chapters/c5718>

Chapter pages in book: (p. 99 - 114)

CHAPTER 4

Recommendations Designed to Reduce Heterogeneity of Data

EVERY STUDY is of course faced with fresh problems, many of which cannot be foreseen; what can be accomplished depends upon the resources that can be tapped and the ingenuity and skill of the investigators. No comprehensive blueprint for all studies can be prepared nor can we outline plans that assure the maximum returns in each case, but in the preceding summary the new data most urgently needed and additional ways in which existing material might be used, as well as the major shortcomings of size distributions of income, are suggested. We now try to implement the suggestions by specific recommendations.

The lack of adequate data for important segments of the field and of exhaustive exploitation of existing data can be remedied through the work of individuals and agencies and through cooperative action. Data from diverse sources, collected for different purposes, can be integrated through cooperative action alone. No matter how well intentioned or able those in charge may be, only by chance will individuals or agencies working independently present their data on comparable bases. The tabulations best suited for one study will not be best suited for integrating the results of several studies. The Conference on Research in Income and Wealth takes advantage of this first comparison of source material on income distribution to lay the groundwork for cooperative action. It seemed the appropriate organization, consisting as it does of representatives from practically all major agencies interested in income data and concerned with their collection, tabulation, and analysis.

The urgent need for coordination is amply demonstrated in the summary. It is almost impossible to find out whether specific differ-

ences between two distributions are attributable to real differences between the groups covered, the dates to which the figures refer, or merely to differences in techniques and concepts. In consequence, much of the value the data would otherwise have is lost. Instead of discovering a missing piece of one jigsaw puzzle, each study turns up a piece of another puzzle.

This unfortunate state of affairs is not, of course, due to deliberate perversity on the part of the agencies responsible for the various bodies of data. And only in small degree does it reflect ignorance of the techniques used by others. In the main, it arises from differences in the purposes for which the data are collected. Seldom is information about income the main objective. For example, the data from the National Health Survey were a by-product of a canvass for information on medical care; from the Study of Consumer Purchases, of an investigation into family expenditures; from the Financial Survey of Urban Housing, of a study of the real estate market; from the Michigan Census, of a survey of unemployment; and so on. The Minnesota field survey is the one source described in this volume in which, from collection through tabulation, income data were the primary objective. And its product, more than any other, can be put into such form that it can be compared with the by-products of other studies.

The purposes for which data are collected naturally determine their character; e.g., what we can learn from income tax returns is in large measure determined by the revenue laws. And as we have seen in Chapter 1, different purposes require different kinds of data on income. Any recommendations designed to facilitate comparability must recognize this situation. It would be neither wise nor practicable to force income data collected for diverse objectives into a single rigid mold.

Nevertheless, the various bodies of data are often sufficiently similar for connecting links to be forged at relatively small cost. To request that the agency responsible for any body of data provide such connecting links seems reasonable, especially since they would be of considerable value to the agency itself, enabling it to use other data to better advantage. Although the purposes for which data from income tax returns are tabulated demand con-

cepts different from those adapted to other purposes, income can usually be calculated from the returns according to concepts applicable to other data as well. It is too much to ask that separate sets of tabulations be made for each concept. But is it too much to ask that among the regular tabulations a table be inserted showing the relation between the concept utilized for tax purposes and one other concept designed to give the maximum comparability with data derived from other sources? Such a table would not bridge all the gaps between income tax and other data, but it would lay a substantial foundation.

Conceived in this spirit, the recommendations concern primarily: (1) differences in the concepts of income adopted; (2) the class intervals by which income is tabulated; (3) differences in coverage; (4) variation in the income unit used, since some studies treat the family as the income receiving unit, others the individual, and the definition of the family varies from study to study; (5) the time and area units. They do not propose uniform procedures; they merely urge a minimum degree of standardization and the inclusion, whenever possible, of a few tabulations designed to facilitate comparability.

I THE CONCEPT OF INCOME

If a single concept of income can be accepted as a basis for comparison, each study could construct one or more tables cross-classifying its own concept with the common or comparison concept.¹ For example, in studies of tax returns such a table would show the distribution of individuals in each net income class by size of income according to the comparison concept (Table 8 is illustrative). Two functions would be served: (1) one set of marginal totals would yield a distribution of income according to the comparison concept that would be comparable from study to study; (2) the table as a whole would provide a basis for estimating the effect on other tables of modifying the concept. Some additional cost would be involved, arising from (a) a breakdown of income items such that income could be computed according to the comparison concept, (b) the making of one or more extra tables. The first cost can be minimized through a careful choice of the concept,

but in some studies it may still be so great as to prevent the adoption of this recommendation, particularly when no more than a rough index of economic status and merely a single global estimate of income are desired (e.g., the National Health Survey,

TABLE 8

Illustrative Form for Table Classifying Income Receiving Units by Income According to Comparison Concept and by Income as Defined in Specific Study

INCOME CLASS (comparison concept)	TOTAL INCOME RECEIVING UNITS	NUMBER OF INCOME RECEIVING UNITS WITH INCOME (AS DEFINED IN SPECIFIC STUDY) OF								
		Less than -\$1,000	-1,000 to -800	-800 to -600	-600 to -400	-400 to -200	-200 to 0	0 to 200	200 to 400	400 to etc.
All classes										
less than -\$1,000										
-1,000 to -800										
-800 to -600										
-600 to -400										
-400 to -200										
-200 to 0										
0 to 200										
200 to 400										
400 to 600										
600 to 800										
800 to 1,000										
1,000 to 1,200										
1,200 to 1,400										
etc.										

Michigan Unemployment Census, and Financial Survey of Urban Housing). In other studies the first cost is unlikely to be crucial. The extra tabulating costs may be relatively great in some instances since an extra computation for each schedule or return may be necessary. However, these costs can be minimized if the tabulations are prepared for in advance or made indirectly.²

In proposing this concept we are not seeking to set up a single theoretically correct all-purpose standard concept of income. How misguided such an attempt would be is obvious even without the reminder in Chapter 1. Since we are concerned here solely with integrating diverse bodies of data, our aim is to select a concept that is applicable at least cost to regularly reported data and that

is likely to be applicable to special studies that may be made. What we are after might perhaps be described as a concept that would be the least common denominator of the data likely to be available. As such it may not be theoretically valid for any specific purpose. It is designed to be the key to a house, not the house itself.

No great sacrifice to expediency is entailed. By means of the cross-classification tables not only can the various studies be integrated, but also the size distributions can be converted into distributions based on various other concepts. Since income tax returns constitute the major source of regularly reported data, considerable weight has been given to the possibility of deriving the concept from federal and state returns.

Bodies of data frequently differ also with respect to the definitions of component items (e.g., wages and salaries). Though raising the same problem, this type of discrepancy cannot be remedied so easily as differences in concepts of income. Most studies furnish data on the principal components of income from which totals can be computed according to alternative concepts, but few furnish data on the parts of the principal components. Consequently, the kind of remedy suggested for differences in income concepts is rarely applicable to components. In the main, we must take the components as we find them. However, in commenting below on the components of the comparison income concept, we call attention to the points at which discrepancies are most likely to occur. Whenever possible, the component item as defined below and as defined in the specific study should be cross-classified. Since important bodies of data are available on salaries and wages, a common basis for this item is imperative.

It is recommended that the comparison income concept be defined as the algebraic sum of the following receipts:

- A Money wages and salaries from nonrelief activity
- B Earnings from relief employment
- C Pensions arising from wage or salary contracts
- D Income from independent business or profession
- E Income from roomers and boarders
- F Interest and dividends

G Rents and royalties

H Income from fiduciaries and trusts

We now comment on the exact definition of these items.³

A MONEY WAGES AND SALARIES FROM NONRELIEF ACTIVITY

Earnings from occupations covered by the Social Security Act should be computed before employee contributions have been deducted, first, because the Social Security wage distributions use this concept; second, because wages are reported in this form on income tax returns; third, because individuals are more likely to report their earnings to enumerators in this form.

The exclusion of income in kind is dictated mainly by difficulties of valuation. Another reason is the wide diversity with which wage and salary payments in kind have been treated. The Census requested information on money wages and salaries alone; the Consumer Purchases Study, on money wages and salaries plus rent as pay; the federal income tax returns, on money wages and salaries plus all wages and salaries in kind except value of housing received by ministers of the gospel, officers and enlisted personnel of the Army, Navy, Coast Guard, Coast and Geodetic Survey, and Public Health Service, and value of living quarters furnished for the convenience of the employer; the Minnesota Income Study, the Social Security Board, and some state income tax returns, on money wages and salaries plus all wages and salaries in kind. Thus the primary data run the whole gamut from the narrowest to the broadest concept. Whenever possible, wages and salaries in kind should be separated from money wages and salaries, and total wages and salaries cross-classified with money wages and salaries.

B EARNINGS FROM RELIEF EMPLOYMENT

Work relief earnings are apparently taxable by the federal government and at least some states. Even were they excluded from taxable income, however, the amount would seldom, if ever, be sufficient to destroy the applicability of the comparison concept to income tax data.

C PENSIONS ARISING FROM WAGE AND SALARY CONTRACTS

Pensions of this type should be included when received. Often a form of deferred wage payment, they are in varying degree included in the recipient's taxable income. Old-age benefits under the Social Security program should be included even though not subject to federal income tax. To the extent that pension and benefit receipts are returns of contributions they may appear under (A) in one time period and (C) in another, but there seems to be no practicable alternative method of treatment.

D INCOME FROM INDEPENDENT BUSINESS OR PROFESSION

This item should be on a net income rather than a withdrawal basis. In computing net income, inventory profits or losses should be estimated as the value of the physical addition to or subtraction from the inventory, rather than as the change in the value of the inventory, i.e., the difference between the value of the inventory at the end and beginning of the year. Automobile expense attributed to the business should be treated like other expenses and deducted in calculating net income. Farm benefit payments should be included in net farm income. Because of wide disparity of treatment and the difficulties of and lack of uniformity in valuation, income in kind should be excluded. Though the federal income tax applies to income in kind of nonfarm enterprises, it is doubtful that a large amount is reported. The instructions specifically exclude the income in kind of farmers. The content of net income from an independent business or profession, like wages and salaries, varies from study to study. Cross-tabulations between income computed according to the concept here defined and according to other concepts (e.g., withdrawals, gross income) would be highly desirable.

E INCOME FROM ROOMERS AND BOARDERS

Though admittedly difficult, an attempt should be made to compute a net income for this item. The chapters on the Minnesota Income Study and the Study of Consumer Purchases describe methods that have been used.

F INTEREST AND DIVIDENDS

Stock dividends, dividends from non-paid up insurance policies, and refunds from cooperatives should be excluded. The first will almost certainly not be reported to field enumerators, and since they can be issued in such a form as to be nontaxable, some will not be entered on returns. Tax exempt interest renders rigorous comparability between income tax and other data with respect to interest impossible. Variation in the issues that are exempt makes for noncomparability even among income tax data for different jurisdictions. The lack of comparability can be somewhat lessened if such tax exempt interest as is reported on tax returns is included. A better solution might be to revise the form so that tax exempt interest would be entered as an item under total income and then subtracted under deductions.

G RENTS AND ROYALTIES

Comment on this item would be superfluous; we merely point out that a net figure is desirable. For practical reasons the federal income tax provisions for depreciation should be followed and the imputed rental value of owner-occupied dwellings excluded because it is not reported for tax purposes. Share rent should be included.

H INCOME FROM FIDUCIARIES AND TRUSTS

These items should be counted in the income of the beneficiary and should consist of income actually distributed. Trust income, however, may be taxable to the beneficiary, trust, or grantor. Precisely how the distribution of net income for tax purposes is affected by including trust income in the beneficiary's income and also by adding undistributed income should be studied.

With the exceptions noted, the comparison concept can be derived from federal tax returns and probably from almost all state tax returns, though both the lack of uniformity among the states and the exclusion of out-of-state income by some states may decrease somewhat the usefulness of distributions for the latter. The items specified are almost certain to be covered in all future studies

that give income in detail rather than a single global estimate. The concept recommended thus satisfies the two major requirements.

The comparison concept cannot, of course, serve as a substitute for concepts on which studies have been based. It excludes for practical reasons certain types of receipt that clearly constitute income, also various types of receipt that for some purposes may be treated as income or tabulated with income but would be excluded from income as defined or tabulated for other purposes. Two examples of items that are clearly income but yet are excluded from the comparison concept are:

- I The value of products produced and consumed at home
- J Imputed income from owner-occupied urban and farm dwellings and from other consumer durable goods used by their owners

Both items are difficult to measure and, with probably minor exceptions, are not entered on tax returns, but any study gathering details on income will doubtless include them so far as possible.

Other types of receipt omitted from the concept but that for some purposes may be considered to be, or tabulated with, income are:

- K Unemployment compensation
- L Benefits, workmen's compensation, accident benefits, etc.
- M Regular contributions received for support
- N Relief and welfare assistance
- O Pensions not arising from wage or salary contracts or from annuities, insurance settlements, and payments under the Old-Age and Survivors Insurance program
- P Inheritances, settlements from insurance not purchased by recipients
- Q Capital gains and losses
- R Prizes, rewards, and gambling gains
- S Gifts other than regular contributions

Items (I) through (N) may be the chief sources of receipts among the lower income groups and have been and probably will be covered in surveys but are not usually reported on tax forms.

The treatment of items (K) through (S), whose theoretical

status is ambiguous, varies widely among tax jurisdictions; furthermore, special studies rarely attempt to separate the part that constitutes a return on capital, if indeed the items are covered at all.

Inheritances and receipts from insurance not purchased by the recipients are not reported on income tax forms. The Consumer Purchases Study included the part of such receipts that was used for family living as income; the Minnesota Study included the entire amount in total income.

The proper treatment of capital gains and losses would be very difficult to decide on theoretical grounds alone. For some purposes this item should clearly be included in the income of individuals; for other purposes, it should no less clearly be excluded, and the intermediate area is broad. Since income tax authorities and various studies treat it differently, it has been excluded from the comparison concept.

Finally, in studies in which income and expenditures are balanced, information on certain non-income receipts should be collected and tabulated:

- T Withdrawals from bank accounts
- U Receipts from other conversions of assets into cash used for living expenses
- V Net receipts from borrowings

These three items are omitted from the definition of income for virtually all purposes and are rarely obtained either from tax returns or studies. Nevertheless, they are needed to balance receipts and expenditures in consumer purchase and other studies.

The income items that will have to be deducted from the concept used by a specific study to attain the comparison concept will vary greatly in importance: one item may occur frequently and be large; others may be picayune. A significant item is likely to be of interest in its own right and its deduction possible at little extra cost through the use of tables that would be prepared in any event. Although deduction of the other items may be troublesome and costly, we trust these recommendations will not therefore be disregarded but will be looked upon as a goal to be striven for. The federal income tax data furnish an illustration. Total income as

reported on federal returns differs from the comparison concept chiefly because it includes capital gains and losses. The deduction of this one item would be a big step toward comparability even if adjustments could not be made for other differences, such as the small amount of income in kind reported. An additional improvement would be to use total income as the basis of classification in some tables.

II INCOME CLASS INTERVALS USED IN TABULATION

The various studies described use a variety of income intervals. The differences among them are so great that preparation of tables comparing data from the various sources is extremely difficult. In all the studies the distributions are based on arithmetic intervals, but only three income classes common to all the data can be obtained from the intervals used: under \$1,000, \$1,000-2,000, and over \$2,000. For most of the data the middle interval can be divided into the two classes, \$1,000-1,500 and \$1,500-2,000. For incomes above \$2,000, a \$2,000-3,000 class can be obtained for all studies except one; and intervals of \$3,000-5,000 and \$5,000 and over can be obtained for the same studies, except for Social Security data. The lack of comparability in the income intervals is especially serious for the range below \$1,000, where so large a part of total income is concentrated.

Very little attention has been given by statisticians and economists to the preparation of standard class intervals, or to the principles that should guide the preparation of standard class intervals of distributions covering as wide a range of values as do income distributions. We cannot ask that all studies tabulate solely by standard class intervals, but much would be gained if at least the same class intervals were used for the comparison concept in the cross-classification table recommended above. In formulating a proposal for standard class intervals for income tabulations, the following principles have been selected as important:

- a) Small intervals are necessary in the lower income ranges because of the very large number of individuals and families involved;

- b) The intervals should be convenient for machine tabulation;
- c) The detailed intervals proposed should be such that combination into a moderate number of intervals (not over ten) covering the entire income range can be prepared on the basis of a common principle.

The first two principles in combination lead us to recommend \$100 or \$200 class intervals for incomes from zero to at least \$1,000, preferably to \$3,000. Adherence to the third principle can be achieved only by selecting intervals built on a geometric, rather than an arithmetic, differential. However, selection of intervals on a precise geometric ratio basis throughout the entire income scale is impossible to reconcile with the use of intervals most convenient for machine tabulation. The most suitable geometric ratio seems to be 2 (or possibly 2.5).

After consideration of various possibilities, we recommend that the intervals given below be used as standard for the classification of income into a moderate number of intervals covering the entire income range. Of the ten intervals all conform to the geometric ratio principle except two; in these there is a slight deviation from the multiple of two. We recommend that all income distribution studies use intervals that can be converted into these standard intervals.

Under 0	\$1,500- 3,000
\$0 - 200	3,000- 6,000
200- 400	6,000-12,000
400- 800	12,000-25,000
800-1,500	25,000 and over

For tabulations of individual earnings and for studies of family income covering the lower income ranges, it is recommended that either \$100 or \$200 class intervals be used up to \$3,000, and that if the \$200 interval is used, an additional break be made at \$1,500. Because of the heavy concentration of recipient units in these ranges it is important to use as many intervals as practicable. It is further recommended that, in general, tabulations be prepared for the income range from \$3,000 to \$10,000 in \$1,000 intervals, and from \$10,000 to \$20,000 in \$2,000 intervals. Above \$20,000,

intervals of \$5,000 or appropriate multiples thereof may be used. We recommend use of the following:

\$25,000- 50,000
50,000-100,000
100,000-250,000
250,000 and over

III COVERAGE OF INCOME

Several of the chief bodies of data provide information solely or primarily on wages and salaries; e.g., the Social Security records, which promise to be the most important regularly reported information for lower income groups, and the data from the 1940 Census, the most extensive in coverage. The task of passing from distributions of wages and salaries to distributions of total income would be greatly facilitated by data on the relation between wages and salaries and total income. It is therefore recommended that all studies that compile data on both total income and wages and salaries cross-classify income receiving units by wages and salaries and total income (Table 9 is an illustration). If possible, the comparison concept should be used, the income class intervals should be convertible into those listed above, and the wage and salary class intervals should be combinations of or convertible into those used in the Census and Social Security tabulations.⁴

Rent or rental value if the dwelling is owner-occupied, can often be ascertained for groups for which income cannot. They would be useful in deriving or testing income distributions if the relation between rent and income were known. Tables cross-classifying rent and income by size, and if possible using the comparison concept and class intervals convertible into those recommended above, are suggested.

IV THE INCOME RECEIVING UNIT

No two studies defined the income receiving unit in exactly the same way. One major source of nonuniformity would be eliminated if data from income tax returns were made internally consistent. This could be largely accomplished by (1) inserting a

question on the form requesting the individual, if married and living with the spouse, to give the total income of the spouse, and (2) preparing a tabulation classifying husband and wife by their joint income. While this procedure would facilitate combining data for joint and separate returns, it by no means gives family income, since members other than the husband and wife may receive income.

TABLE 9

Illustrative Form for Table Classifying Income Receiving Units by Income According to Comparison Concept and by Earnings

INCOME CLASS (comparison concept)	TOTAL INCOME RECEIVING UNITS	NUMBER OF INCOME RECEIVING UNITS WITH EARNINGS OF					etc.
		No earnings	Less than \$100	100 to 200	200 to 300	300 to 400	
All classes							
less than -\$1,000							
-1,000 to -800							
-800 to -600							
-600 to -400							
-400 to -200							
-200 to 0							
0 to 200							
200 to 400							
400 to 600							
600 to 800							
800 to 1,000							
1,000 to 1,200							
etc.							

There seems no feasible way of making bodies of data from sources that define the income receiving unit in different ways more comparable, but experimentation should not cease. Cost will ordinarily rule out the type of solution recommended to meet the analogous problem of differences in income concept. Studies that use a family unit should attempt to make it as nearly like that used in other studies as possible. They will probably not always have means to present additional tabulations of individual income,

but they should do so whenever they can. Although the Committee has pondered how to cross-classify individual and wage income with family income, it has reached no generally satisfactory solution. Wage distributions would be of considerably greater value if they could be converted into distributions of family income; Social Security data, for example, would thereby become far more useful. But no key to conversion has been found, and no method of devising one invented.

V THE TIME AND AREA UNITS

Almost all annual data on income or related items are on a calendar year basis. Consequently integration is difficult if a non-calendar year is used, as in the Study of Consumer Purchases and the Minnesota Field Survey. As in these two studies, circumstances may render the use of a calendar year virtually impossible. When it is not, a calendar year is recommended.

Likewise almost all data relate to the distribution of annual incomes and provide information for each unit for one year only.⁵ To understand the distribution of income adequately such data must be supplemented by information for a longer period. The value of any study will be greatly enhanced if it can obtain data on the income of the income receiving unit in each of several other years. Accumulation of data of this type would pave the way for a thorough study of the stability of income status, of the extent to which differences in annual incomes misrepresent income differentials for a longer period.

For some purposes the value of the data will be increased substantially if they are tabulated on the basis of relatively small geographical units although expense and adequacy of coverage will obviously limit the detail that can be expected. Economic areas do not necessarily coincide with jurisdictions established by law, but since many data will be most easily procured on a state basis, and since they will always pertain to some state and county, it seems desirable to recommend tabulation by states and counties. Tabulation by size of community also will often be needed, but we are not able to recommend any specific size grouping.

NOTES

¹ An appropriate name for this concept, device, or measure fails us. Connotations attaching to 'uniform', 'standard', or 'basic' make them unsuitable. Since the objective is to construct a device that will facilitate the integration of different sets of data, but that will not necessarily be used itself, 'conversion concept', 'conversion measure', or 'integration concept' might be preferable to 'common concept'.

² For example, if only one or two items need be deducted from the concept used in the study to attain the comparison concept, the table could be obtained indirectly by cross-classifying the former concept by the item or items to be excluded. An extra computation for each return would be avoided. Also, when punch cards are used, the cards containing the items in question could be sorted out. The computations could then be made solely for the cards containing the items.

³ The importance of precise, comprehensive definitions can scarcely be over-emphasized. The student attempting to use the materials in Part II may encounter a good many difficulties because the terms were not carefully enough defined.

⁴ The most detailed wage or salary class intervals being used by the Census in tabulating individual incomes are: \$0; \$100 intervals to \$2,999; \$500 intervals to \$4,999; \$5,000 and over; unknown. Less detail will be shown in the family tabulations and in most of the individual tabulations.

The Social Security and Old-Age Insurance data for 1937 are tabulated by \$100 intervals from \$0 to \$3,000, and have an over \$3,000 class; the data for 1938 are tabulated by \$200 intervals to \$3,000 with one class for \$3,000 and over.

⁵ Several bodies of data discussed in Part II constitute exceptions, notably the Wisconsin Income Study, the Delaware Study, and the Financial Survey of Urban Housing.