STANDARDIZED COMPARISONS IN POPULATION RESEARCH*

EVELYN M. KITAGAWA
Population Research and Training Center, University of Chicago

RESUMEN

La estandarización es usada en el análisis demográfico para hacer comparaciones de la incidencia o prevalencia de varios fenómenos, tomados en cuenta con las diferencias en uno o más factores componentes. Este artículo se propone explorar la amplitud de las alternativias disponibles para las composiciones controladas y, especialmente, las relaciones de la estandarización con las teorías de la construcción de números índices desarrolladas por los economistas. No sólo tasas de incidencia, sino también aquellas medidas de prevalencia tales como promedios, razones y distribuciones porcentuales pueden ser estandarizadas mediante una variedad de alternativas, estando determinada la elección de la mejor de éstas por la naturaleza de la comparación a ser hecha. La amplitud de los procedimientos de estandarización puede ser clasificada dentro de dos clases generales, "Agregativos" y "promedio de relativos." Estandarizaciones directas e indirectas, como las usadas por los demógrafos, son subtipos de índices agregativos, en tanto que varias de las medidas especiales de mortalidad que han sido desarrolladas, son subtipos de la clase promedios de relativas. Este examen sistemático de la estandarización puede tener dos usos en la metodología demográfica:

a) *Enfatiza que las medidas estandarizadas son índices comparativos de un procedimiento, y no medidas del procedimiento mismo; una tasa estandarizada, por ejemplo, es altamente útil para propósitos analíticos, pero no tiene valor intrínseco en sí misma.*

b) *Se dispone de varios sistemas alternativos de construcción de números índices para computar medidas estandarizadas, y algunos de ellos (tal como el Indice Ideal de Fisher) merecen ser más explorados por los demógrafos.*

Comparisons of population phenomena are often complicated by related factors which should be taken into account in any definitive analysis. The need to control for differences in population composition, when comparing rates of incidence of some phenomenon in two or more populations, is generally recognized. It is taken for granted, for example, that differences between crude death rates of various communities do not provide sufficient grounds for ranking the communities in terms of longevity and health conditions. Because the size of the crude death rate (deaths per 1,000 total population) is closely related to age composition, two communities with identical schedules of age-specific death rates may have very different crude death rates if one has a much higher proportion of retired persons than the other. For this reason mortality analyses typically rely on comparisons of schedules of age-specific death rates for the various population groups involved. However, as the number of populations to be compared increases, the number of differences between age-specific rates to be examined and interpreted becomes difficult to manage, especially when the death rates are also specific for such factors as sex, race, and sometimes occupation or socio-economic status. In these circumstances the need for summary measures which compare the incidence of mortality independent of differences in age composition, and possibly other compositional factors, is readily apparent.

The technique of standardization is commonly used to make such "controlled" comparisons of the incidence of various phenomena, with differences in one or more compositional factors taken into account. However, to this writer's knowledge, the full range of alternatives for controlling composition that are available to the researcher—and the rationale underlying each—have not been system-

* The view of standardization set forth in this paper is based in large part on work done in the course of preparing a chapter on standardization for a "Manual on Techniques of Demographic Research." The writer is greatly indebted to Donald J. Bogue, senior author of the forthcoming manual, for his critical review of several versions of the chapter, and to Otis Dudley Duncan, whose comments on the first draft were most helpful.

atically elaborated in expositions of demographic techniques. Discussions of methods for making mortality comparisons provide the most definitive statements available in the demographic literature concerning alternative standardization procedures. This substantive orientation, however, tends to limit consideration to those standardized measures that have been utilized in mortality analysis and these primarily in terms of their utility and validity for mortality comparisons.[1] Jaffe's summary of statistical methods for standardization of populations presents a broader perspective and includes, in addition to direct and indirect standardization of rates, a discussion of the so-called "Westergaard method of expected cases" and "expectancy (or probability) tables."[2] In a previous article the writer systematically explored the relationship between "differences in general (unstandardized) rates" and "differences in standardized rates."[3]

Relatively little attention has been directed, in the literature on demographic methods, to other standardized comparisons, including the possibilities for averages and entire distributions of population characteristics, although the argument for controlled comparisons applies to these aspects of population research as well. For example, differences in the occupational distribution of two groups are not unrelated to differences in their educational composition and it may be desirable, in some analyses, to make the occupational comparison with differences in education level taken into account. Similarly, differences in the average income of two groups may be due to differences in their occupational composition, and it may be desired to compare their income levels independent of occupational composition (i.e., with occupation controlled).

The purpose of this article is to present a systematic and generalized summary of standardization techniques in population research—covering averages and entire distributions of population characteristics as well as the more commonly discussed standardized rates and indexes. The terminology in discussions of standardized measures is somewhat confusing and ambiguous. Particularly unfortunate in some respects is the use of the term "indirect standardization," which at times refers to rates intended as approximations to standardized rates computed by the direct method, at other times to ratios of actual to expected numbers which are computed and interpreted from the standpoint of their own "direct" meanings, and at still other times to a standardized index which purposely has a different set of weights from one population to another when computed for a series of populations. To avoid further confounding the situation, the discussion below starts from some conventional definitions and then proceeds to "rename" them in a systematic framework.

STANDARDIZED RATES

The term *rate* is used here to refer to the ratio of the "number of occurrences of an event" to the "population exposed to the event." That is, $c = e/n$, where e = number of events, n = exposed popula-

[1] See, for example, Theodore D. Woolsey, "Adjusted Death Rates and Other Indices of Mortality," chap. iv in *Vital Statistics Rates in the United States*, 1900–1940, by Forrest E. Linder and Robert D. Grove (Washington: Government Printing Office, 1943); Peter R. Cox, *Demography*, chap. vii (New York: Cambridge University Press, 1950); J. Yerushalmy, "A Mortality Index for Use in Place of the Age-adjusted Death Rate," *American Journal of Public Health*, XLI (August, 1951), 907–22; H. Silcock, "The Comparison of Occupational Mortality Rates," *Population Studies*, XIII (November, 1959), 183–92; S. J. Kilpatrick, "Occupational Mortality Indices," *Population Studies*, XVI (November, 1962), 175–87; Hugh H. Wolfenden, "On the Theoretical and Practical Considerations Underlying the Direct and Indirect Standardization of Death Rates," *Population Studies*, XVI (November, 1962), 188–90.

[2] A. J. Jaffe, *Handbook of Statistical Methods for Demographers*, chap. iii (Washington: Government Printing Office, 1951).

[3] Evelyn M. Kitagawa, "Components of a Difference between Two Rates," *Journal of the American Statistical Association*, L (December, 1955), 1168–74.

tion, and c = rate. Conversely, the number of events may be thought of as the result of the operation of the rate in the population, that is, $e = c \times n$. Specific rates refer to the rate of occurrence of an event in subgroups of a population when classified by some trait or characteristic (for example, age), while a general rate refers to the rate for the total population that is so classified. Consequently, the general rate for a particular population may be written as:

$$c = \frac{\Sigma n_i c_i}{n_.} = \Sigma \frac{n_i}{n} c_i, \qquad (1)$$

where

$c_.$ = general rate for the total population;

c_i = the specific rate for persons in the ith category of trait I in the population (i.e., c_i represents the schedule of I-specific rates);

n_i = the number of persons in the ith category of trait I in the population (i.e., n_i defines the I-composition);

$n_.$ = total number of persons in the population ($n_. = \Sigma n_i$).

In other words, the general rate in a population may be viewed as the result of the operation of its schedule of I-specific rates (c_i) on its own I-composition (n_i or $n_i/n_.$). For example, the general (or crude) death rate of a population may be obtained simply by dividing the number of deaths by the total population exposed to the risk of death,[4] but it may also be interpreted as a weighted average of the age-specific death rates using the age composition as the set of weights.

Standardized rates (*direct method*).—The use of a standardized rate controls composition by applying a schedule of specific rates to a corresponding schedule of compositional categories of a population that has been adopted as a standard for a particular comparison. (Standardized rates are sometimes referred to as "adjusted" rates, especially in mortality analysis.) The I-standardized rate for a given population is defined as follows:

$$I\text{-std. rate (direct)} = \frac{\Sigma N_i c_i}{N} \quad \text{or} \quad \Sigma \frac{N_i}{N_.} c_i, \qquad (2)$$

where

N_i = number of persons in ith category of trait I in the standard population;

$N_.$ = total number of persons in the standard population;

c_i = the specific rate for persons in the ith category of trait I in the given population.

Trait I in this definition refers to the population trait or characteristic over which the standardization is made, for example, age, if the rate is being standardized to control for age composition. From equation (2) it is clear that the I-standardized rate of a given population is a weighted average of its I-specific rates (c_i), with the I-composition of the standard popluation ($N_i/N_.$) as the weights.

If the standardization controls for two population characteristics simultaneously —say traits I and J—the result of the multiple standardization may be written as:

IJ-std. rate (direct)

$$= \frac{\Sigma_i \Sigma_j N_{ij} c_{ij}}{N} \quad \text{or} \quad \Sigma \frac{N_{ij}}{N} c_{ij},$$

where the ij subscript represents the ijth subcategory in the cross-classification by traits I and J simultaneously. For purposes of simplification the discussion in this article is limited to standardization for one trait only (I). However, it can readily be extended to two or more traits simply by treating the IJ cross-classification—or the IJK cross-classification, if

[4] The crude death rate is usually expressed as a rate per 1,000 population, but the algebra and discussion of standardization are simplified if rates are left as ratios without multiplication by 1,000 or some other constant, and this practice is followed here.

three characteristics are controlled—as a single distribution cross-classified by two (or three) characteristics simultaneously. Also, to simplify the algebraic notation the Σ sign will be used without a subscript to indicate summation over the categories of trait I.

Interpretation of standardized rates.—A standardized rate has no intrinsic value in itself. The numerical value of an age-standardized death rate, literally interpreted, indicates what the crude death rate would be in the "hypothetical" population which has the age-specific death rates of the given population and the age composition of the standard population. A series of age-standardized death rates (for a series of "given" populations) based on the same standard population, therefore, indicates what the crude death rate would be in a series of different hypothetical populations, each characterized by the schedule of age-specific death rates for one of the given populations and the age composition of the standard population. The purpose of standardization, however, is not to determine the absolute level of the crude death rate in a succession of hypothetical populations but rather to effect *comparisons* of the schedules of age-specific death rates for the series of given populations with *differences in age composition controlled.* From this perspective the type of control over composition that is accomplished by direct standardization may be described as follows: The effect of composition is not eliminated; it is merely held constant according to an arbitrarily adopted schedule of weights which defines a "standard" composition and to which one or more schedules of specific rates are applied. Most of the problems that arise in making standardized comparisons stem from the fact that the choice of a standard population can have a significant influence on the results obtained and is especially difficult when the populations being compared have very marked compositional differences with respect to the factors controlled in the standardization.

Indirect standardization of rates.—When schedules of specific rates are not available for the populations to be compared but data on their compositional structure can be obtained, the following formula is often used to compute a standardized rate by the so-called indirect method:

$$I\text{-std. rate (indirect)} = \left(\frac{e.}{\Sigma n_i C_i}\right) C. = \left(\frac{c.}{\Sigma \frac{n_i}{n.} C_i}\right) C., \qquad (3)$$

where

$\frac{n_i}{n.}$ defines the I-composition of the given population;

$c. = \frac{e.}{n} =$ the general (unstandardized) rate in the given population, or the number of events divided by the total population;

$C. =$ general rate in the standard population;

C_i defines the I-specific rates in the standard population.

This computation requires both the general rate and the I-specific rates for the standard population, but only the general rate and the I-composition of the series of given populations to be compared. The indirect rate for each given population is obtained by multiplying the general rate for the standard population by an adjustment factor

$$\left(\frac{e.}{\Sigma n_i C_i}\right)$$

which is equal to the ratio of "actual" to "expected" events for the given population, where the expected number represents the number of events that would have occurred in the given population if it experienced the I-specific rates of the standard population while retaining its own I-composition. For example, the adjustment factor for an indirectly age-standardized death rate is the ratio of the actual number of deaths that occurred in the given population to the number that would have occurred if the age structure of the given population were exposed to the age-specific death rates of the standard population. Inasmuch as the adjustment factor is the only variable in a series of indirectly standardized rates based on the same standard population, it can be used "as is" to make standardized comparisons by the indirect method. There is no logical reason for converting a series of adjustment factors to "rates" by multiplying them by the constant, $C.$, since it has already been pointed out that the absolute numerical values of standardized rates are of little interest.

The following algebraic equivalents for the adjustment factor indicate the indirect type of control over composition that is accomplished by indirect standardization:

$$\frac{e.}{\Sigma n_i C_i} = \frac{c.}{\Sigma \frac{n_i}{n.} C_i} = \frac{\Sigma \frac{n_i}{n} c_i}{\Sigma \frac{n_i}{n} C_i}.$$

The last expression shows that the adjustment factor is equal to the ratio of the "directly standardized rate for the given population" to the "directly standardized rate for the standard population," using the composition of the given population as the set of weights for both computations. That is, the adjustment factor for a given population is, in reality, a summary index which directly compares the schedule of specific rates for the given population with the schedule of specific rates for the so-called standard population, using the composition of the *given* population as the *standard* composition for the comparison. Accurate interpretation of comparisons standardized by the indirect method, therefore, must take into account two inferences which follow from the above: (1) the so-called standard population does not provide the set of compositional weights used in indirect standardization; instead, its role may more accurately be described as that of a "base population" for a standardized "index" computed by the indirect method; (2) strictly speaking, composition is not held constant in a series of indirectly standardized rates or indexes because the set of weights implicit in the standardized measure for each "given" population is its own composition and this varies from one population to the next. The implications of these characteristics of indirect standardization will be expounded more fully in the discussion of standardized indexes.

The label, "indirect standardization," probably stems from the use of this technique as a substitute for direct standardization, although the fact that it only "indirectly" controls for compositional differences may also be relevant. Indirect standardization may be used as a substitute for direct standardization for any of the following reasons: (1) the indirect method involves fewer computations; (2) unlike direct standardization, it does not require schedules of specific rates for all the populations involved in the comparison; (3) its results may be more stable (that is, have smaller sampling errors) than those of direct standardization and, consequently, can be used when numbers run too small in the subgroups for which specific rates must be computed in direct standardization. Justification of indirect standardization in these terms, however, implicitly assumes that directly standardized rates would be preferred if the data and facilities for their computation were available and if there were no major problems of statistical reliability in the schedules of specific rates to be compared. However, procedures that are equivalent to indirect standardization can and have been justified as the preferred method of making controlled comparisons given particular research questions and objectives. These, too, are discussed in the section on standardized indexes which follows.

STANDARDIZED INDEXES

When standardization techniques are used to make comparisons of rates holding selected compositional factors constant, there are several advantages to expressing the results as index numbers. First, a standardized index is, by definition, a comparative summary measure. That is, it directly summarizes the differences between two schedules of specific rates. Second, the implications of alternative sets of weights (standard compositions) for a particular analysis are clarified when indexes are used. Finally, when the results of standardization are expressed as index numbers, it is at once apparent that there is a formal identity between standardization in population research and the construction of index numbers in economics. This opens up a framework for standardization that is broader in scope than the computation and comparison of standardized rates. (For example, the second general form of a standardized index discussed below is not derived from standardized rates.) Also, standardized comparisons in demography have many of the complications and ambiguities involved in the comparison of price indexes and, consequently, the economists' penetrating discussion of these problems can be applied with benefit to demographic comparisons. Needless to say, the economists' evaluation of alternative indexes cannot be transplanted intact to demographic comparisons, which deal with basic data that differ in certain important respects from economic data and which include geographical and compositional comparisons as well as time series.

A standardized index may take one of two general forms: (1) a ratio of one weighted average (or sum) to another weighted average (or sum) or (2) a weighted average of ratios or "relatives." In order to construct a series of index numbers which compare the schedules of specific rates for a series of populations, it is necessary to select both a standard population, whose compositional strata will be used to control composition, and a base population, whose schedule of specific rates will be used as the base for each index number. Each index is a relative number in which the specific rates for one of the populations being compared appear in the numerator and the specific rates for the base population appear in the denominator.

An index number of the first general form mentioned above—an aggregative index in the language of the economist—is defined as the ratio of the "weighted average (or sum) of the given population's schedule of specific rates using the composition of the standard population as weights" to the "weighted average (or sum) of the base population's schedule of rates using the same standard composition as the weights." This definition is expressed algebraically in the following formula:

I-std. index (aggregative)

$$= \frac{\Sigma N_i c_i}{\Sigma N_i c_i'} = \frac{\Sigma \dfrac{N_i}{N.} c_i}{\Sigma \dfrac{N_i}{N.} c_i'}, \quad (4)$$

where

c_i represents the schedule of I-specific rates of the given population for which the standardized index is computed;

N_i or $\dfrac{N_i}{N}$ represents the I-composition of the standard population;

c_i' represents the schedule of I-specific rates of the base population.

Similarly, an index number of the second general form—an "average of relatives" in economic terminology—may be defined as a weighted average of "specific ratios," that is, of the ratios of the specific rates for the given population to the corresponding rates for the base population. The general formula for an "average of relatives" index is:

I-std. index (relatives)

$$= \frac{\Sigma N_i \dfrac{c_i}{c_i'}}{N.} = \Sigma \left(\frac{N_i}{N.}\right)\left(\frac{c_i}{c_i'}\right). \quad (5)$$

The difference between the two general forms of index numbers may be summarized as follows: the aggregative index first averages (or aggregates) each schedule of specific rates by applying the schedule of rates to the standard composition and then takes the ratio of the resulting averages (or aggregates); the average of relatives first computes the ratios of corresponding specific rates and then averages the resulting schedule of specific ratios. It is clear from equation (4) that an aggregative index, from the perspective of population research, is also equivalent to the ratio of two standardized rates. Specifically, it is equal to the ratio of the standardized rate for the given population to the standardized rate for the base population, using the composition of the standard population as the weights.

The economic literature on price indexes critically evaluates numerous alternatives of both forms of index numbers.[5] The various mortality indexes discussed in the demographic literature (see n. 1) have their counterparts in the economists' system of index numbers and are identified in the discussion that follows.

Aggregative index, base population as standard.—If the base population is chosen as the standard population, the following aggregative index is obtained:

$$I\text{-std. index (aggreg. \#1)} = \frac{\Sigma n_i' c_i}{\Sigma n_i' c_i'} = \frac{\Sigma n_i' c_i}{e_.'} = \frac{\Sigma \frac{n_i'}{n_.'} c_i}{\Sigma \frac{n_i'}{n_.'} c_i'} = \frac{\Sigma \frac{n_i'}{n_.'} c_i}{c_.'}, \qquad (6)$$

where c_i refers to the I-specific rates of the given population; $e_.'$, $c_.'$ and c_i' refer to the total number of events, the general (unstandardized) rate, and the schedule of I-specific rates, respectively, in the base population; and $n_i'/n_.'$ refers to the I-composition of the base population. The various algebraic equivalents reported in equation (6) indicate that this index is equal to the following ratios, any of which might be used as a definition: (1) the ratio of "expected events" to "actual events" in the base population, with expected events defined under the assumption that the I-specific rates of the given population operate on the I-composition of the base population; (2) the ratio of the "I-standardized rate for the given population" to the "I-standardized rate for the base population" using the I-composition of the base population as the standard; and (3) the ratio of the same "I-standardized rate for the given population" to the "general rate of the base population."

Economists call this form of an aggregative price index the Laspeyres Index. Demographers refer to the mortality index computed by this method as the Comparative Mortality Figure. This index is equivalent to the direct method of standardization of rates discussed earlier.

Aggregative index, given population as standard.—If the I-composition of the given population is used as the standard for a series of comparisons, the following index is obtained:

$$I\text{-std. index (aggreg. \#2)} = \frac{\Sigma n_i c_i}{\Sigma n_i c_i'} = \frac{e_.}{\Sigma n_i c_i'} = \frac{\Sigma \frac{n_i}{n_.} c_i}{\Sigma \frac{n_i}{n_.} c_i'} = \frac{c_.}{\Sigma \frac{n_i}{n_.} c_i'}. \qquad (7)$$

The algebraic equivalents shown in this equation lead to the following alternative definitions: (1) the ratio of "actual events" to "expected events" in the given population, with expected events defined under the assumption that the I-specific rates of the base population operate on the I-composition of the given population; (2) the ratio of the "I-standardized rate for the given population" to the "I-standardized rate for the base population" using the I-composition of the given population as the standard; and (3) the ratio of the "general rate for the given population" to the same "I-standardized rate for the base population."

[5] For example, see Irving Fisher, *The Making of Index Numbers*, 3d ed. (Boston and New York: Houghton Mifflin Company, 1927); and F. E. Croxton and D. J. Cowden, *Applied General Statistics*, 2d ed. (New York: Prentice-Hall, 1955), chap. xvii and xviii.

Economists call this form of an aggregative price index the Paasche Index. The mortality index computed by this method is known as the Standardized Mortality Ratio. This index is also equivalent to the so-called indirect method of standardization, which was discussed earlier in this paper, and to Westergaard's "method of expected cases."

Aggregative index, average population as standard.—In this index the average composition of the given and the base population is used as the standard composition for the comparison, as shown in the formula below:

I-std. index (aggreg. #3)

$$=\frac{\Sigma\left(\frac{n_i}{n_.}+\frac{n_i'}{n_.'}\right)c_i}{\Sigma\left(\frac{n_i}{n}+\frac{n_i'}{n'}\right)c_i'}=\frac{\Sigma\frac{1}{2}\left(\frac{n_i}{n_.}+\frac{n_1'}{n_.'}\right)c_i}{\Sigma\frac{1}{2}\left(\frac{n_i}{n}+\frac{n_i'}{n'}\right)c_i'}. \quad (8)$$

The last algebraic expression indicates that this index is equivalent to the ratio of the "I-standardized rate for the given population" to the "I-standardized rate for the base population" using their average composition as the standard. An aggregative price index of this type is called the Marshall-Edgeworth Index by economists, and the equivalent mortality measure is known as the Comparative Mortality Index in the demographic literature.

Another variant of equation (8) provides the simplest formula for purposes of computation.

I-std. index (aggreg. #3)

$$=\frac{c+\Sigma\frac{n_i'}{n_.'}c_i}{c'+\Sigma\frac{n_i}{n_.}c_i'} \quad \text{(8 alternate)}$$

To compute this index, it is necessary to obtain two sum products, one ($\Sigma n_i'c_i$) by applying the I-composition of the base population to the I-specific rates of the given population, and the second ($\Sigma n_i c_i'$) by applying the I-composition of the given population to the I-specific rates of the base population.

Simple aggregative index.—This index is defined as follows:

I-std. index (aggreg. #4)

$$=\frac{\Sigma c_i}{\Sigma c_i'}=\frac{\Sigma\frac{c_i}{k}}{\Sigma\frac{c_i'}{k}}, \quad (9)$$

where k = number of categories of trait I, the trait for which the index is standardized. The simple aggregative index is the ratio of "the sum (or arithmetic mean) of the I-specific rates of the given population" to the "sum (or arithmetic mean) of the I-specific rates of the base population." In effect, this index assigns equal weight to each category of the classification of trait I.

If trait I is age or some other quantitative variable, and the intervals by which it is classified are not of equal size—for example, some 5-year and some 10-year age intervals—the simple aggregative index is adjusted for variation in size of interval as follows:

$$I\text{-std. index (aggreg. \#4)}=\frac{\Sigma a_i c_i}{\Sigma a_i c_i'}, \quad (10)$$

where a_i = size of ith interval of trait I (e.g., number of years in ith age interval). The gross reproduction rate is a simple aggregative rate commonly used in demographic analysis. It is defined as the sum of the age-specific birth rates of women in the child-bearing ages, and the ratio of two gross reproduction rates is, in effect, a simple aggregative index. In mortality analysis, the simple aggregative index is known as the "ratio of equivalent average death rates."[6]

Aggregative index, other population as standard.—Rather than selecting one of the standard compositions described

[6] See Woolsey, *op. cit.*, p. 81, and Cox, *op. cit.*, p. 116, for a discussion of problems of age coverage in this mortality index.

above, it may be desired to use the composition of some other population—real or theoretical—as the standard for a particular analysis. In this case, the standard composition selected can be substituted for N_i in equation (4), the general formula for an aggregative index.

Averages of relatives.—Instead of first aggregating each schedule of specific rates and then taking the ratio of two aggregates, index numbers may also be computed by first computing individual ratios of corresponding specific rates in two schedules—c_i/c_i' in the notation used here—and then averaging these ratios. Simple and weighted averages of various types—arithmetic means, harmonic means, and geometric means—have been used by economists in the construction of price indexes.

The *simple* average (arithmetic) of relatives is defined as:

I-std. index (relative #1)

$$= \frac{\Sigma \frac{c_i}{c_i'}}{k} \quad \text{or} \quad \frac{\Sigma a_i \frac{c_i}{c_i'}}{\Sigma a_i}. \qquad (11)$$

In the first formula, k equals the number of intervals into which trait I is classified; if trait I is a quantitative variable and its intervals are unequal in size, the second formula in equation (11) is used, in which a_i equals the size of the ith interval of trait I. This "unweighted" average of relatives, in effect, assigns equal weight to the ratio, c_i/c_i for each category of trait I (or each unit of a quantitative variable), and it has been recommended as a substitute for the age-adjusted death rate (direct method) by Yerushalmy (see n. 1).

A *weighted* average (arithmetic) of relatives has the following general formula:

$$I\text{-std. index (relative \#2)} = \frac{\Sigma N_i \frac{c_i}{c_i'}}{N_.}, \qquad (12)$$

where N_i = weight for the ith category of trait I. Any of the standard compositions described for aggregative indexes may be used in this formula. When the composition of the given population (n_i) is used as weights, equation (12) reduces as follows:

I-std. index (relative #3)

$$= \frac{\Sigma n_i \frac{c_i}{c_i'}}{n_.} = \frac{\Sigma \frac{e_i}{c_i'}}{n_.}. \qquad (13)$$

In mortality analysis, this is called a Relative Mortality Index. It has the advantage that the only information needed for the given population is its total size ($n_.$) and the age composition of deaths (e_i) and it, therefore, can be computed for intercensal years when statistics on deaths are compiled but data on the age composition of many populations are not available.

Alternative sets of weights for various averages of price relatives—including geometric and harmonic means as well as arithmetic averages—have been evaluated by economists, in terms of criteria relative to time series analyses of prices. While not all of their discussions are directly relevant to the problem of making standardized comparisons in population research, knowledge of the characteristics of the numerous forms which index numbers can take will facilitate the demographer's evaluation of alternative indexes for his own analyses.

Fisher's ideal index.—Two tests have been used by economists to measure the bias inherent in various index numbers. The time reversal test requires that an index for year t with year 0 as base must be the reciprocal of an index for year 0 with year t as base. The factor reversal test requires that, if both a price index and a quantity index are computed by the same formula for the same period (that is, the same given year and base year), their product must be equal to the ratio of "total values in the given year" to "total values in the base year." None of the indexes described above meets both tests. After examining a variety of formulas designed to "rectify" the biases inherent in different indexes by averaging geometrically formulas which err in opposite directions, Fisher selected one of those which satisfied both tests as "ideal" from the standpoint of accuracy and ease of computation.[7] This ideal index is the geometric mean of the Laspeyres and Paasche aggregative indexes. In the notation adopted here for population research, the ideal index would be defined as:

$$\text{Ideal index} = \sqrt{\frac{\Sigma n_i' c_i}{\Sigma n_i' c_i'} \times \frac{\Sigma n_i c_i}{\Sigma n_i c_i'}} = \sqrt{\frac{\Sigma n_i' c_i}{\Sigma n_i c_i'} \times \frac{e.}{e.'}}. \quad (14)$$

One implication of the factor reversal test for population research may be of special interest. This test is equivalent to requiring that, if "an index summarizing differences in I-specific rates of two populations" is multiplied by "the same index summarizing their differences in I-composition," the product must be equal to the ratio of the general rates of the two populations. That is, the factor reversal test requires that an index number separate the proportionate difference between two general rates ($c./c.'$) into two component indexes, one summarizing differences in I-specific rates and the other differences in I-composition.

Theoretical comparison of alternative indexes.—The manner in which an aggregative index summarizes differences between two schedules of specific rates is clarified by subtracting "1" from equation (4), thereby converting it from a "ratio of two weighted averages" to a measure of the "proportionate difference between them," as follows:

$$I\text{-std. index (aggreg.)} - 1 = \frac{\Sigma \frac{N_i}{N.}(c_i - c_i')}{\Sigma \frac{N_i}{N.} c_i'}. \quad (15)$$

This transformation makes it clear that *an aggregative index, in effect, summarizes absolute differences between corresponding specific rates in two schedules of rates.* That is, in this form the aggregative index is a weighted average of the absolute differences between corresponding specific rates ($c_i - c_i'$), expressed as a proportion of the weighted average of the base schedule of rates.

However, if "1" is subtracted from equation (5), the general formula for an average of relatives index, the following result is obtained:

$$I\text{-std. index (relatives)} - 1 = \Sigma \frac{N_i}{N.}\left(\frac{c_i - c_i'}{c_i'}\right). \quad (16)$$

In this form, *an average of relatives index is shown to be a weighted average of proportionate or relative differences between corresponding specific rates in two schedules of rates*

$$\left(\frac{c_i - c_i'}{c_i'}\right).$$

Thus, the choice between an aggregative index and an average of relatives for a particular comparison requires a decision as to whether one wishes to summarize *absolute* or *relative* differences between the schedules of specific rates. Yerushalmy has argued that in mortality comparisons one is primarily interested in comparative risks and, therefore, in relative differences in age-specific rates, and he proposes the simple average of relatives as an age-adjusted mortality index.[8] On the other

[7] Irving Fisher, *op. cit.*, chap. xi.

[8] *Op. cit.* The reader of Yerushalmy's article may not immediately recognize his proposed index as a simple average of relatives, however, because his computation procedure focuses attention on the derivation of a set of weights that can be applied directly to the schedules of specific rates for the series of populations to be compared. In our notation Yerushalmy's index may be written as:

$$\Sigma\left(\frac{100}{\Sigma a_i} \times \frac{a_i}{c_i'}\right) c_i,$$

where the factor in parentheses represents the set of weights to be applied to each schedule of specific rates (c_i). It is easily demonstrated that this formula is equivalent to the simple average of relatives reported in equation (11), and Yerushalmy's justification and interpretation of the hand, the Registrar-General of England and Wales has selected various aggregative indexes for mortality comparisons, with some preference for the Comparative Mortality Index (average composition as standard) to analyze changes in mortality over time, and the Standardized Mortality Ratio (composition of given population as standard) to compare occupational differentials in mortality.[9] Unfortunately, the choice between an aggregative index and an average of relatives is not a simple one, and it is not independent of the selection of a logical set of weights. For example, the general formula for an aggregative index can be converted to an average of relatives as follows:

I-std. index (aggreg.)

$$=\frac{\Sigma N_i c_i}{\Sigma N_i c_i'}=\frac{\Sigma (N_i c_i')\dfrac{c_i}{c_i'}}{\Sigma N_i c_i'}. \qquad (17)$$

That is, an aggregative index is equivalent to an average of relatives with the I-composition of "expected events" ($N_i c_i'$) as the weights. While this relationship may explain why comparisons based on aggregative indexes Nos. 1 and 2, for example, can differ so markedly from comparisons based on average of relatives No. 3, it does not help much in the selection of a set of weights. It has been assumed throughout this discussion that, in population research, the set of weights for both types of index will typically be the composition of some population or perhaps equal weights as in the simple average of relatives and the gross reproduction rate. Presumably, these are logical weights for our purpose,[10] and the choice between an aggregative index and an average of relatives should be made on the basis of whether one wants to summarize absolute or relative differences in specific rates, ignoring the fact that the aggregative index is equivalent to an average of relatives with the composition of events as weights.

The implications of this choice may be clarified by emphasizing one of the definitions for the general aggregative index given in equation (4), namely, the ratio of $\Sigma N_i c_i$ to $\Sigma N_i c_i'$. That is, an aggregative index compares directly the *numbers of events* resulting from two schedules of specific rates, with composition held constant. From this perspective, *aggregative indexes may be interpreted as comparisons of standardized numbers of events.* Accordingly, the choice between an aggregative index and an average of relatives in a mortality analysis, for example, should be made on the basis of whether the researcher wants to compare two schedules of death rates in terms of the total number of deaths they would yield in a standard population *or* in terms of the relative (proportionate) differences between corresponding specific rates in the two schedules. Both types of index can be useful when correctly applied and interpreted.

It must be recognized at the outset, however, that no single summary statistic can be a substitute for a detailed comparison of the specific rates in two or more schedules of rates. Each of the indexes defined in this chapter is a summary measure, and when the differences between corresponding specific rates vary markedly from one category to another of trait I—in some cases, they may even have opposite signs—a single summary average of the specific differences may be misleading. For instance, an index of 1.00 indicates that the average difference between two schedules of rates is zero using a specified set of weights. But such a result can be obtained from widely different situations, for example, when all the specific differences are zero *or* when substantial differences of opposite sign cancel each

index are in terms of a measure designed to summarize relative differences in age-specific rates by assigning equal weights to every age.

[9] Cox, *op. cit.*, pp. 119–32.

[10] This is not the case in price indexes, where quantities are used as weights for aggregative indexes but cannot be used as weights for averages of relatives because the units in which quantities are reported are not comparable for different items of goods included in the index. In population research, however, the units of population—people—do not present a comparability problem.

other when weighted by the standard composition selected for the comparison. Similarly, indexes close to 1.00 may mask wide differences of opposite sign between specific rates. This consideration has led researchers to caution against the use of single summary measures such as standardized rates or indexes without supplementing such measures with analyses of differences between the specific rates. However, the same persons have recognized the definite need for a single figure which summarizes the numerous differences between two or more sets of specific rates.

The selection of a particular index—once the general form has been decided—requires specification of a base population and a standard composition. If only two populations are involved in the comparison, then one can be used as the base population and their average composition may provide a reasonable set of weights, although in a particular analysis another set of weights may answer a question that is more relevant to the research objectives. When the comparison involves three or more populations, the situation is more complicated. If all comparisons can meaningfully be made with one (fixed) population, the latter is a logical choice for the base population. If, however, the various populations are to be compared directly with each other, the selection of a base population is more difficult, and it also must be recognized that in some indexes the standard composition changes from one index number to another in a series. Among the indexes described in this paper the following do not, strictly speaking, hold composition constant from one index to another in a series: (1) aggregative index, given population as the standard; (2) aggregative index, average of given and base population as the standard; (3) average of relatives, given population as the standard. This does not imply, however, these indexes are necessarily inferior. Economists often recommend the second (Marshall-Edgeworth) index for price comparisons over time, and note has already been taken of the use of both the first and second indexes by the Registrar-General of England and Wales for mortality comparisons.

The dilemma of "direct" versus "indirect or variable" control over composition has been discussed by a number of writers. Yule contended that those indexes characterized by "indirect or variable" control over composition (that is, those for which the set of weights changes from one index number to another in the series being compared) are not fully methods of standardization at all, but are safe only for the comparison of single pairs of populations, namely, comparison of each given population with the base population.[11] Stocks, however, in discussing Yule's statement, took the position that the indirect method of the Standardized Mortality Ratio (aggregative index, given population as standard) seems

> to suffer no greater disadvantage, either theoretical or practical, than the direct method of obtaining such a ratio [aggregative index, base population as standard] provided that we understand clearly that by the indirect method a good or a bad mortality at a given age has a greater importance attached to it in a group of people where that age is well represented than in another group where it is poorly represented,—whereas by the direct method equal importance is attached to a mortality at a given age whether people of that age form 10 or 50 percent of the population concerned. Which of these is the more desirable seems to me a matter of opinion.[12]

Cox also acknowledged the usefulness of the Standardized Mortality Ratio, noting that it has the "merit that it continually raises the question of the comparability of the age-distributions" of the populations whose rates are being compared, and sum-

[11] G. Udny Yule, "On Some Points Relating to Vital Statistics, More Especially Statistics of Occupation Mortality," *Journal of Royal Statistical Society*, XCVII, Part 1, (1934), 1–72.

[12] "Discussion" on Mr. Yule's paper, *ibid.*, pp. 75–76.

marizes the dilemma in the following terms:

> The need to use the same set of weights in a series of index numbers is incompatible with the inappropriateness of the same set of weights if the age distributions differ substantially. This difficulty is inherent in the nature of index numbers themselves and cannot be completely resolved. There can thus be no such thing as a perfect mortality index. Any concession towards one or another of these two incompatible considerations depends on the nature of the populations being compared in any particular case.[13]

In addition to the dilemma of "direct" versus "variable" control over composition is the problem of selecting a base population for a comparison. Even those indexes which exercise direct control over composition are subject to possible complications in the selection of a base population when the researcher compares the index numbers for various populations with each other. For example, the simple average of relatives directly controls composition in the sense that the same standard composition (equal weights) is used throughout a series, but comparison of the indexes for any pair of populations in the series does not necessarily give the same result that would be obtained if one of the two populations were used as the base for this particular comparison (economists call this the "Circular Test"). For example, comparison of Yerushalmy's mortality index for two different states, with the United States as the base population, would not necessarily give the same result as computing his index for one state using the other state as the base population. Among the indexes discussed in this paper only two, the aggregative index with base population as standard (which is equivalent to direct standardization of rates) and the simple aggregative index, have the characteristic that any pair of indexes in a series can be compared with each other with the same result that would be obtained if one of the two populations involved were used as the base population.

[13] Cox, *op. cit.*, pp. 117–21.

Each of the various indexes discussed in this paper, in effect, answers a particular question about differences between schedules of specific rates, and the researcher must decide which question to ask. For instance, the "aggregative index with given population as standard" compares the actual number of events in the given population with the number that would occur if it were exposed to the specific rates of the base population. In some situations this may be a meaningful way to summarize differences between schedules of rates. For example, in comparing the death rates of different socio-economic groups in a population, the highest socio-economic group might be chosen as the base population and the composition of each "given population" used as the set of weights in computing its own mortality index. Such an index—computed for each socio-economic group in turn—would measure the proportion of "excess deaths" in each socio-economic group under the assumption that its death rates could be as low as those for the highest socio-economic group if it had access to equivalent living conditions, medical facilities, etc. It is, however, up to the researcher to decide if this is the most meaningful way for him to summarize socio-economic differentials in mortality given the particular objectives of his analysis.

Another factor to be taken into account in selecting an index is the availability of data. If specific rates are available for a base population but not for all the other populations whose schedules of rates are to be compared, only two of the indexes described above are feasible: (1) the aggregative index with given population as standard, which requires only the general rate and the I-composition of the given populations; and (2) the average of relatives with the composition of given population as weights, which requires only the I-composition of events and the total size of the given populations. It should be mentioned, however, that when these indexes are used without any knowledge of the pattern of differences between corresponding specific rates, the conclusions are subject to the hazards of single summary measures described earlier.

Demographers have never, to this writer's knowledge, used an equivalent of Fisher's ideal price index to make standardized comparisons in population research. However, as was mentioned earlier, this index does satisfy both the "time reversal" and "factor reversal" tests and it also can be used to allocate the "relative" difference between two general rates ($c_./c_.'$) into two component indexes—one summarizing differences in specific rates and the other differences in composition. The relationship between this type of allocation into components and that discussed in a previous paper[14] may be worth pursuing. In the previous paper it was *the absolute difference between two general rates that was allocated into the sum of two components,* as follows:

$$c_. - c_.' = \Sigma \tfrac{1}{2}\left(\frac{n_i}{n} + \frac{n_i'}{n_.'}\right)(c_i - c_i') + \Sigma \tfrac{1}{2}(c_i + c_i')\left(\frac{n_i}{n_.} - \frac{n_i'}{n_.'}\right). \qquad (18)$$

The first component is a weighted average of the absolute differences between corresponding specific rates ($c_i - c_i'$) of two populations, using their average composition as the weights; the second is a weighted average of absolute differences between the I-composition ($n_i/n_. - n_i'/n_.'$) of the two populations, using an average of their I-specific rates as weights.

Fisher's ideal index could similarly be used to allocate *the relative difference between two general rates into two component indexes,* as follows:

$$\frac{c_.}{c_.'} = \underbrace{\sqrt{\frac{\Sigma n_i' c_i}{\Sigma n_i' c_i'} \times \frac{\Sigma n_i c_i}{\Sigma n_i c_i'}}}_{\text{component index of rates}} \times \underbrace{\sqrt{\frac{\Sigma c_i' \dfrac{n_i}{n_.}}{\Sigma c_i' \dfrac{n_i'}{n_.'}} \times \frac{\Sigma c_i \dfrac{n_i}{n_.}}{\Sigma c_i \dfrac{n_i'}{n_.'}}}}_{\text{component index of composition}}. \qquad (19)$$

The first component index in this equation is the ideal index shown in equation (14), which is the geometric mean of two aggregative indexes summarizing differences in I-specific rates—one using the I-composition of the base population as weights, and the second the I-composition of the given population as weights. Similarly, the second component index in equation (19) is the geometric mean of two indexes summarizing differences in I-composition; one an aggregative index using the I-specific rates of the base population as weights, and the second an aggregative index using the I-specific rates of the given population as weights. It is easily demonstrated that the product of these two component indexes is equal to $c_./c_.'$, the ratio of the general rates of the two populations. For purposes of computation the following equivalents may be used to advantage:

Component index of rates

$$= \sqrt{\frac{\Sigma n_i' c_i}{\Sigma n_i c_i'} \times \frac{e_.}{e_.'}},$$

Component index of composition =

$$= \sqrt{\frac{n_.' c_.}{n_. c_.'} \times \frac{\Sigma n_i c_i'}{\Sigma n_i' c_i}}.$$

Example of results.—The results of some illustrative computations are reported in Table 1, utilizing death statistics for Massachusetts, Mississippi, and the United States in 1950. The need to control for differences in age composition in this mortality comparison is dramatically indicated by the change in *direction* of the differ-

[14] Kitagawa, *op. cit.*

ences in mortality when age is controlled. That is, although the crude death rate for Mississippi was one percent below, and the rate for Massachusetts 9 percent above, the crude death rate for the entire country (9.64 deaths per 1,000 population), all of the age-standardized indexes are higher for Mississippi than Massachusetts. The various indexes, however, do not all provide the same evaluation of the over-all *size* of mortality differences. In general, the averages of relatives show much larger "average" differences in mortality than the aggregative indexes, but the choice of a standard composition has relatively little effect on the comparison, especially if the range of choice is limited to the actual compositions of the populations whose rates are being compared. For example, when the United States is used as the base population and the composition of the given population provides the weights, the aggregative index (Standardized Mortality Ratio in Table 1) indicates that Mississippi's age-specific death rates are *in general* 14 percent higher than the age-specific rates for the United States, and 21 percent higher than the age-specific rates for Massachusetts. However, when the same composition is used as weights in an average of relatives (Relative Mortality Index in Table 1), age-specific death rates in Mississipi are in general 38 percent higher than in the United States and 72 per cent higher than in Massachusetts. Very similar results are obtained if the composition of the base population is used as the set of weights.

Table 1.—AGGREGATIVE INDEXES, AVERAGES OF RELATIVES, AND AGE-SPECIFIC DEATH RATES PER 1,000 POPULATION: MASSACHUSETTS AND MISSISSIPPI, 1950

(U.S. Population as Base)

Age	Age-Specific Death Rates (per 1,000 population)			Absolute Differences		Relative Differences	
	U.S. (1)	Mass. (2)	Miss. (3)	Mass. & U.S. (2)-(1)	Miss. & U.S. (3)-(1)	Mass. & U.S. [(2)÷(1)]-1	Miss. & U.S. [(3)÷(1)]-1
Under 1	32.99	26.88	43.18	-6.11	+10.19	-.185	+.309
1 - 4	1.39	1.03	2.14	-0.36	+0.75	-.259	+.540
5 - 14	0.60	0.40	0.79	-0.20	+0.19	-.333	+.317
15 - 24	1.28	0.80	1.91	-0.48	+0.63	-.375	+.492
25 - 34	1.79	1.29	3.08	-0.50	+1.29	-.279	+.721
35 - 44	3.59	2.92	4.96	-0.67	+1.37	-.187	+.382
45 - 54	8.54	7.94	10.35	-0.60	+1.81	-.070	+.212
55 - 64	19.12	18.64	20.81	-0.48	+1.69	-.025	+.088
65 - 74	40.68	41.69	41.93	+1.01	+1.25	+.025	+.031
75 - 84	93.31	90.79	91.31	-2.52	-2.00	-.027	-.022
85 & over	201.97	185.65	199.94	-16.32	-2.03	-.081	-.010

Type of Rate or Index	Name of Mortality Measure	Mass.	Miss.	% diff.*
General Rate	Crude death rate	10.53	9.54	-9
Aggregative Indexes				
#1-Base pop. as standard	Comparative Mortality Figure	.94	1.12	+19
#2-Given pop. as standard	Standardized Mortality Ratio	.94	1.14	+21
#3-Average pop. as standard	Comparative Mortality Index	.94	1.13	+20
#4-Simple, ages 0-64	Indexes based on	.91	1.21	+33
#4-Simple, ages 65-84	equivalent average	.99	.99	0
#4-Simple, ages 0-84	death rates	.97	1.04	+ 7
Averages of Relatives				
#1-Simple	Yerushalmy's Mortality Index	.84	1.26	+50
-Base pop. as standard	...	.78	1.38	+77
#3-Given pop. as standard	Relative Mortality Index	.80	1.38	+72

*Mississippi minus Massachusetts, expressed as percent of Massachusetts.

These results are not unusual in mortality comparisons based on all causes of death for entire communities. Different patterns may emerge, however, in analyses of particular causes of death or in comparisons of various subgroups (for example, occupational groups) of a population. Also, it is not unusual, when comparing two schedules of death rates, to find that one population has higher rates at some ages but lower rates at other ages. In such a situation aggregative indexes and averages of relatives may well give opposite conclusions as to which population's over-all level of mortality is higher. At any rate it is clear that given the general pattern of age differentials in mortality and the general "pyramid" shape of the age composition of most populations, the choice between an aggregative index or an average of relatives—that is, the choice between summarizing absolute or relative differences in age-specific death rates—can have an important effect on conclusions about the over-all size of mortality differentials.

Interpreted in another way, the aggregative indexes in Table 1 show that the age-specific death rates for Mississippi would result in 19 to 21 percent more deaths than would the death rates for Massachusetts if it is assumed that both sets of rates operate in a population with an age composition within the general range of the actual age distributions of the two states. However, the averages of relatives in Table 1 show that individual age-specific death rates in Massachusetts are, on the average, 72 to 77 percent higher than corresponding age-specific rates in Mississippi when the weights used fall within the general range of the actual age distributions of the two states and 50 percent higher if equal weights are assigned to each age (Yerushalmy's index). The fact that the aggregative indexes give much smaller differences in mortality is explained by the pattern of "absolute" and "relative" differences in age-specific rates reported in the last four columns in the upper half of Table 1. The absolute differences are much smaller in the younger ages (which have large weights when any "real" age composition is the standard) than in the older ages (which have very low weights in real age compositions), whereas the "relative differences" are much larger in the younger ages than in the older ages. This difference in the age pattern of absolute and relative differences is typical of mortality comparisons, since the very low incidence of mortality in the younger ages inevitably tends to produce small absolute but large relative differences at these ages. In a particular analysis, the objectives of the research may favor the use of one or the other general type of index, or the researcher may even wish to compute both an aggregative index and an average of relatives in order to summarize both absolute and relative differences in specific rates.

STANDARDIZATION OF MEANS AND OTHER AVERAGES

An arithmetic mean may be standardized for one or more compositional characteristics by treating means of subgroups of the population—where the subgroups are categories of the characteristic for which standardization is desired—as a schedule of specific rates. Following the direct method of standardization, the standardized mean is a weighted average of the subgroup means using a selected standard composition—population distribution by subgroups—as the weights. For example, in comparing the average income of migrants and non-migrants, the researcher may wish to control for differences in age composition. The age-standardized mean income of migrants is obtained as follows: (1) compute the mean income of each age subgroup of migrants; (2) apply these age-specific mean incomes to the age composition of some population adopted as standard; (3) sum the products of the age-group means and the numbers in the corresponding age groups of the standard population in order to obtain a "total expected income"; and (4) divide the "total expected income" by the total number of persons in the standard population to determine the age-standardized mean income. If the age composition of non-migrants is used as the standard composition in this computation, the resulting standardized mean income of migrants can be compared directly with the unstandardized mean income of non-migrants. If some other age composition is used as the standard, then an age-standardized mean income must also be computed for non-migrants by a similar procedure.

The arithmetic mean may also be standardized by the indirect method, and standardized indexes summarizing differences between two or more sets of subgroup means may be computed by the same procedures used to compare two or more schedules of specific rates. However, these techniques cannot be used to standardize medians and modes. The algebra of standardized rates and indexes applies only to measures which when multiplied by the number of persons in a population yields an expected number of events (or expected total amount of some variable) for the population. Rates, percents, ratios, and arithmetic means satisfy this requirement, but the median, for example, is a "positional" average and does not. Measures such as the median and mode of a distribution can be standardized by first standardizing the entire distribution from which they are computed, as described in the section that follows, and then computing the median or mode of the standardized distribution. The same procedure can also be used to standardize measures of dispersion and other summary measures computed from frequency distributions.

STANDARDIZATION OF PERCENT DISTRIBUTIONS

When comparing percent distributions for two or more populations, it may be desired to standardize the entire distributions for a compositional characteristic which is related to the distributions being compared. This can be accomplished by treating each percent in the distribution as a rate. For example, to standardize the occupational distribution of two or more populations for differences in age composition, each population's percent distribution of workers by occupation is treated as a set of unstandardized percents. If the direct method of standardization is used, the age-standardized occupation distribution for each population is computed by applying its schedule of age-specific percents in each occupation to the age composition of the standard population, summing the results for each occupation and dividing this sum by the total number in the standard population. The schedule of age-specific percents in each occupation expresses the number of workers of each age in that occupation as a percent of the corresponding age total of the population. For example, the age-standardized percent of professional workers in population A is obtained by: (1) computing a schedule of age-specific percents for professional workers, which express the number of professional workers of each age in population A as a percent of all workers of that age in population A; (2) multiplying this schedule of age-specific percents by the age composition of the standard population (and dividing by 100 to adjust for the use of percents instead of proportions), summing the results and expressing the sum as a percent of the total number of workers in the standard population. If the same procedure is followed for each of the other occupation categories, an age-standardized percent distribution by occupation is obtained for population A. This standardized distribution will add to 100 percent (within the limits of rounding).

The formula for computing an age-standardized percent distribution by the direct method may be written as:[15]

$$I\text{-std. proportion in } k\text{th subgroup of pop. (direct method)} = \frac{\Sigma_i N_{i.} \dfrac{n_{ik}}{n_{i.}}}{N_{..}}, \qquad (20)$$

[15] In order to simplify the formulas, the distribution of the given population is expressed algebraically as "proportions" which should be multiplied by 100 to convert to percents.

where

I = population trait for which the proportion in the kth subgroup of the given population is standardized;

$\frac{n_{ik}}{n_{i.}}$ = proportion of persons in the ith category of trait I who are also in kth subgroup of trait K in the given population (or I-specific proportions for subgroup k of given population);

$N_{i.}$ = number of persons in the ith category of trait I in the standard population (or I-composition of the standard population).

Two or more standardized percent distributions (based on the same standard composition) may be compared by any of the procedures used to compare regular (unstandardized) distributions. A particularly convenient summary measure for this purpose is the index of dissimilarity, which is equal to one-half the sum of the absolute values of the differences between corresponding percents in two distributions, and which may be interpreted as the minimum proportion of either population which must be redistributed among the subgroups of the distribution in order for the two populations to have identical percent distributions.

Although the standardization of percent distributions could be a useful analytical device, it has seldom been used in demographic and social research. This is probably due in part to the fact that this application of standardization techniques is not elaborated in the literature, to this writer's knowledge. It is also true that, in many situations where it might be used to advantage, the necessary data are not available to standardize distributions by the direct method described above. For instance, in the example mentioned earlier age-specific percents for each occupation may not be available for all of the populations to be compared. However, if age-specific percents are available for a suitable "standard population," then an age-standardized occupational distribution can be computed by an indirect method which requires only the over-all age composition and the over-all occupational composition of the given populations. Indirect standardization of percent distributions requires a special formula, however, because equation (3) which was used for indirect standardization of rates will not produce a standardized percent distribution that necessarily adds to 100 per cent.[16] Instead, an alternative formula for indirect standardization may be used:[17]

I-std. proportion in kth subgroup of pop. (indirect method)

$$= \frac{N_{.k}}{N_{..}} + \left(\frac{n_{.k}}{n_{..}} - \frac{\Sigma n_i \frac{N_{ik}}{N_{i.}}}{n_{..}} \right), \tag{21}$$

where

$\frac{n_{.k}}{n_{..}}$ = the unstandardized (general) proportion in the kth subgroup of the given population (e.g., the proportion of the given population who are in the kth occupation);

$\frac{N_{.k}}{N_{..}}$ = the unstandardized proportion in the kth subgroup of the standard population (e.g., the proportion of the standard population who are in the kth occupation);

$\frac{N_{ik}}{N_{i.}}$ = the I-specific proportion for the kth subgroup of the standard population

[16] Tests of equation (3) in cases where the I-specific percents and the I-composition of the standard population differed greatly from those of the given population yielded standardized distributions which added to as much as 114 percent.

[17] This formula is discussed in Samuel A. Stouffer, "Standardization of Rates When Specific Rates Are Unknown," in A. J. Jaffe, *Handbook of Statistical Methods for Demographers*, pp. 65–70. If standardized proportions computed by this formula are summed over k, the sum is equal to 1, as follows:

$$\Sigma_k \frac{N_{.k}}{N} + \Sigma_k \frac{n_{.k}}{n_{..}} - \left(\Sigma_i \frac{n_{i.}}{n_{..}} \right) \left(\Sigma_k \frac{N_{ik}}{N_{i.}} \right)$$

$$= 1 + 1 - \Sigma \frac{n_{i.}}{n_{.}} = 2 - 1 = 1 .$$

(e.g., the proportion of the ith age group of the standard population who are in the kth occupation);

$n_{i.}$ = the total number in the ith category of trait I in the given population (e.g., the number of persons in the ith age group of the given population).

By this formula, the indirect I-standardized percent in the kth subgroup of the given population is obtained by *adding* an adjustment factor to the unstandardized percent in the kth subgroup of the standard population; this adjustment factor (which is enclosed in parentheses in the formula) is the difference between the actual percent of the given population in the kth subgroup and the "expected percent" in the kth subgroup, where the expected percent is computed by assuming the I-specific percents for the kth subgroup of the standard population prevail in a population with the I-composition of the given population.[18]

As mentioned earlier, any average (or other summary measure) which can be computed from a frequency distribution may be standardized for one or more compositional characteristics simply by standardizing the entire distribution for this characteristic and then computing the average (or other measure) from the standardized distribution. An arithmetic mean computed from a distribution standardized by the direct method is algebraically identical to a mean standardized by the use of equation (2), the formula for direct standardization of rates.

OTHER FORMS OF STANDARDIZATION

This discussion has purposely been limited to standardized comparisons in which the objective is to control—either directly or indirectly—for compositional differences among the populations involved in the comparison. More complex forms of standardization have been devised to accomplish somewhat different objectives. A variety of expectancy (or probability) tables, for example, attempts to answer the following kinds of questions: (1) how many years, on the average, can a person expect to live? (2) what is the probability that a woman will marry in the course of her lifetime? (3) how many female children, on the average, will a cohort of women bear in the course of their lifetime?

The first question is answered by a life table, in which the value of $\mathring{e}_0$ indicates the average length of life of a cohort who are exposed throughout their lives to a given schedule of age-specific death rates. The reciprocal of $\mathring{e}_0$ is called the life table death rate and, although it is equivalent to a weighted average of the schedule of age-specific death rates on which the life table is based, the weights implicit in the average are not designed to control composition. In fact, the set of weights implicit in each life table death rate is itself a function of the schedule of age-specific death rates on which the life table is based and, consequently, the larger the differences between two schedules of age-specific death rates the greater the differences between the sets of weights implicitly assigned to them in deriving life table death rates.

The probability that a woman will eventually marry is contingent not only on the probability that she will marry between successive ages but also on the probability that she will survive to each age, and nuptiality tables are therefore derived by applying age-specific "first marriage" rates to life table populations. Here again, the set of weights used (a life table population) in summarizing a schedule of age-specific marriage rates does not control age composition but rather takes account of mortality. This is characteristic of expectancy tables, and no attempt is made

[18] The formula given in equation (3) for indirect standardization of rates differs from this formula of Stouffer's in that equation (3) *multiplies* the general rate of the standard population by an adjustment factor which may be defined as the ratio (relative difference) of actual to expected rates for the given population.

here to review these standardization techniques.

Another application of standardization not pursued in this paper is the analysis of differences between two or more general rates into components reflecting differences in composition, on the one hand, and differences in specific rates, on the other hand. The potentialities of Fisher's ideal index for such a components analysis were pointed out, however. Two approaches to this problem are elaborated in publications by the writer and by Duncan, Cuzzort, and Duncan.[19]

[19] Kitagawa, *op. cit.*; Otis Dudley Duncan, Ray P. Cuzzort, and Beverly Duncan, *Statistical Geography* (The Free Press of Glencoe, Ill., 1961), pp. 120–28.