6. Standardization of rates and ratios*

Concepts and basic methods for deriving measures that are comparable across populations that differ in age and other demographic variables.

Overview

Epidemiologists are always mindful of <u>population diversity</u>. Virtually every large population is heterogeneous in regard to sociodemographic (e.g., age, gender, education, religion), geographic, genetic, occupational, dietary, medical history, and innumerable other personal attributes and environmental factors related to health. A population can be viewed as a composite of diverse subgroups (ultimately, subgroups of size one, i.e., individuals, but epidemiologic measures break down at that point). Any overall measure or statistic reflects the value of that measure for each of the subgroups comprising the population.

An overall measure that does not take explicit account of the composition of the population is called **crude**. Its value will be an average of the values for the individual subgroups, weighted by their relative sizes. The larger the subgroup, the more influence it will have on the crude measure (i.e., "democracy"). Thus, the death rate for a population is a weighted average of the death rates for its component subgroups. Suppose we consider a population of size N as consisting of five age groups, or **strata**. Each age **stratum** will have a specific number of people, say n_i (i=1 to 5). During the following year, each stratum will experience some number of deaths, say d_i. The total population size, N, is therefore Σn_i, the total number of deaths, D, is Σd_i, and the crude mortality rate is D/N, which can also be written as a weighted average of the **stratum-specific mortality rates**, d_i/n_i, as follows:

$$\frac{D}{N} = \frac{\Sigma d_i}{N} = \frac{\Sigma n_i (d_i/n_i)}{N} = \Sigma(n_i/N)(d_i/n_i) = \Sigma w_i(d_i/n_i)$$

where w_i are the weights (note that $\Sigma w_i = \Sigma(n_i/N) = (\Sigma n_i)/N = \Sigma n_i/\Sigma n_i = 1$).

The crude rate is the simplest and most straightforward summary of the population experience. But mortality is strongly related to age, so the stratum-specific mortality rates will differ greatly from one another. The summary provided by the crude rate glosses over this heterogeneity of stratum-specific mortality rates.

_____

* (An earlier version of the chapter was prepared by Timothy Wilcosky, Ph.D.)

This issue is particularly relevant when we compare rates across populations or time periods, because if the populations differ in composition, then at least some of what we observe may be attributable to these differences. For example, suppose you and a friend each agree to bring 10 pieces of fruit to a picnic. You stop at a fruit stand and buy 8 mangoes ($1.00 apiece) and 2 apples ($0.50 apiece). Meanwhile your friend goes to the supermarket and buys 2 mangoes ($1.75 apiece) and 8 apples ($0.45 apiece). Which is the more expensive purchase? From one perspective, the first purchase is the more expensive, since $9.00 is certainly greater than $7.10. But from another perspective, the second purchase is more expensive, since the supermarket charged a much higher price for the mangoes and only slightly less for the apples.

Which of these perspectives you choose depends on the purpose of your question. More often than not, the epidemiologist (and the serious shopper) would ask whether the prices were higher in the fruit stand or the store and by how much. We can answer that question by simply comparing the price lists. But what if you also bought oranges, melons, grapes, and bananas? What if you bought two dozen varieties of fruit? It would certainly be more convenient to have a summary measure that permitted an overall comparison. The trouble with total cost ($9.00 versus $7.10) or average price ($0.90/piece of fruit versus $0.71 per piece) is that the fruit stand average price gives more weight to the price of mangoes, because you bought more mangoes, whereas the supermarket average price gives more weight to the price of apples because your friend bought more apples. We're comparing apples to mangoes, instead of fruit stand to supermarket.

Clearly what we need is a procedure that averages the prices in the same way for each vendor, so that both averages give the same proportionate weighting to mangoes. The average prices will depend upon the weighting we use, but at least we will be comparing (proportionally speaking) apples with apples and mangoes with mangoes. However, it's also clear that at least in this example, the weights will determine which seller is favored by the comparison. The fruit stand owner will prefer a higher weight for the price of mangoes, so that her prices will seem the better bargain. But the supermarket owner will prefer a very low weight on the mangoes. He might argue, in fact, that mangoes are a specialty item and not really worth considering in the comparison. He might argue for assigning zero weight to the mangoes, so that his average price will be 0.45/piece (the summary is simply the price of the apples), which is less than the fruit stand charges for apples.

Which set of weights is the right one to use? People who don't like mangoes might agree with the supermarket owner. People who like mangoes – or fruit stands – would not. For the most part, the choice of weights (a.k.a. the **standard population**) is based on convention, the intended and potential comparisons, and various other considerations. There is often no absolute correct choice, and there can easily be different opinions about the best one. But it helps to have a rationale for the choice other than that it happens to give you a result you like. Finally, nothing you do about weights is going to change the fact that your purchase **did** cost more than your friend's, so the crude summaries are not irrelevant.

Adjustment and standardization

The terms "adjustment" and "standardization" both refer to procedures for facilitating the comparison of summary measures across groups. Such comparisons are often complicated by

differences between the groups in factors that influence the measures of interest but which are not the focus of attention. Adjustment attempts to remove the effects of such "extraneous" factors that might prevent a "fair" comparison.

"Adjustment", the more general term, encompasses both standardization and other procedures for removing the effects of factors that distort or *confound* a comparison. Standardization refers to methods of adjustment based on weighted averages in which the weights are chosen to provide an "appropriate" basis for the comparison (i.e., a "standard"), generally the number of persons in various strata of one of the populations in the comparison, an aggregate of these populations, or some external relevant population. Other kinds of adjustment, some of which also employ weighted averages, will be discussed in the chapter on **Confounding**.

Most textbooks of epidemiology present the topic of rate standardization in relation to adjusting for age. This tendency is not coincidental, since virtually all mortal or morbid events occur with different frequencies among groups of different ages. But the same principles and procedures apply to subgroups defined by other variables. The following example illustrates how these varying frequencies can affect a summary measure. Table 1 indicates that in 1970, 5,022 out of the 562,887 white women in Miami died, and that 285 of the 106,917 white Alaskan women died. The respective overall (crude) death rates are 8.92 per 1,000 and 2.67 per 1,000. Is life in Alaska more conducive to longevity than life in Florida?

Although the crude rates suggest that the force of mortality is stronger in Miami than in Alaska, Table 1 reveals that for any given age the two populations have very similar mortality rates. What then accounts for the difference in the crude death rates? A look at the age distributions in Miami and Alaska provides the answer. Compared to Alaska, Miami has a much greater proportion of women in older age groups, where mortality is high. Since the data from larger strata dominate the crude death rate, the Miami death rate is heavily influenced by the high mortality in older ages. In contrast, in Alaska the crude death rate reflects the low mortality rates among young women, who account for a much larger proportion of the Alaska population than they do of the Florida population.

Two populations may have the same overall size and identical age-specific death rates, but different total numbers of deaths and different overall death rates, due to differences in their age distributions. Standardization (and other adjustment procedures) seeks to provide numbers and comparisons that minimize the influence of age and/or other extraneous factors.

Table 1
Population and Deaths by Age in 1970 for White Females in
Miami, Alaska, and the U.S.

Age	Miami Pop.	Miami Deaths	Miami Rate*	Alaska Pop.	Alaska Deaths	Alaska Rate*	U.S. Pop.+	U.S. Deaths+	U.S. Rate*
< 15	114,350	136	1.19	37,164	59	1.59	23,961	32	1.34
15-24	80,259	57	0.71	20,036	18	0.90	15,420	9	0.58
25-44	133,440	208	1.56	32,693	37	1.13	21,353	30	1.40
45-64	142,670	1,016	7.12	14,947	90	6.02	19,609	140	7.14
65+	92,168	3,605	39.11	2,077	81	39.00	10,685	529	49.51
	562,887	5,022		106,917	285		91,028	740	
Crude death rate*			8.92			2.67			8.13

* Deaths per 1,000 population   + in thousands

Standardization of rates by the direct method

In the above example, the difference in crude death rates between Alaska and Miami results from differences in their respective age distributions rather than differential age-specific death rates. It follows intuitively that if Miami had the same age distribution as Alaska, or vice-versa, their crude death rates would be similar to each other. As a matter of fact, if Miami and Alaska had the same age distribution, regardless of what that distribution might be, their crude death rates would be similar, since their age-specific rates are similar.

In direct standardization the stratum-specific rates of study populations are applied to the age distribution of a standard population. (In the above example, each age group is a stratum.) Consequently, if Alaska happened to have the same age distribution of white females as the 1970 U.S. white female population, and Miami also had this same age distribution, then the crude death rates for Alaska and Miami would be similar. In other words, direct standardization applies the same set of weights to the age-specific rates of Alaska and Miami, and the summary (age-adjusted) death rate is therefore independent of differences in the age distribution of the two populations. The directly age-standardized death rates are equivalent to the crude death rates which Miami and Alaska "would have experienced" if they had had the same age distribution as the 1970 U.S. white female population.

Computationally, direct standardization of rates is straightforward:

$$\text{Directly standardized rate} \quad = \quad \frac{\sum (\text{stratum-specific rates} \times \text{standard weights})}{\sum (\text{standard weights})}$$

$$\text{Directly standardized rate} \quad = \quad \frac{(r_1 N_1 + r_2 N_2 + r_3 N_3 + \ldots + r_n N_n)}{(N_1 + N_2 + N_3 + \ldots + N_n)}$$

$$R_s \quad = \quad \frac{\sum (r_k \times N_k)}{\Sigma (N_k)} \quad = \quad \sum \left(r_k \times \frac{N_k}{\Sigma (N_k)} \right) \quad = \quad \sum \left(r_k \times \frac{N_k}{N} \right)$$

$$R_s \quad = \quad \sum (r_k W_k)$$

where:

r_k = rate in k-th stratum of the *study* population

N_k = number of persons in k-th stratum of the *standard* population

N = <u>total</u> number of persons in the *standard* population (ΣN_k)

W_k = weight for each stratum (equal to N_k/N)

$\sum$ means summation over the k strata.

This formula shows that, when the same standard is used, if two study populations have the same age-specific rates (i.e., for each k their R_k's are equal) then their directly standardized rates will be identical, independent of the age distributions in the study populations. The standardized death rate for white Miami women using the 1970 U.S. population of white women as the standard is:

$$\text{Directly standardized rate} \quad = \quad \frac{(1.19 \times 23{,}961) + (0.71 \times 15{,}420) + \ldots + (39.11 \times 10{,}685)}{91{,}208}$$

$$= \quad 6.92 \text{ deaths/thousand}$$

The corresponding standardized rate for Alaska is:

$$\text{Directly standardized rate} \quad = \quad \frac{(1.59 \times 23{,}961) + (0.90 \times 15{,}420) + \ldots + (39.00 \times 10{,}685)}{91{,}208}$$

$$= \quad 6.71 \text{ deaths/thousand}$$

(Results can be expressed as decimal fractions or scaled to aid in their intuitive meaningfulness, e.g., 0.00134 = 1.34 per thousand = 134 per hundred thousand.)

After adjusting for age, the difference in death rates between Alaska and Miami is nearly eliminated.

Some points to consider

There are several things to consider about the above formula and computation. First, the directly standardized rate is a <u>weighted average</u>. Since each W_k is the <u>proportion</u> that the k-th stratum is of the total standard population, the weights are simply the proportional age distribution in the standard population. The <u>crude death rate</u> in a population, which represents the total number of deaths divided by the total number of persons, can be regarded as an <u>average of the population's stratum-specific death rates (R_k) weighted by its own age distribution</u>.

Similarly, a directly standardized rate corresponds to the crude rate that would be observed in the standard population if the standard population had the same stratum-specific rates as does the study population. (To put the foregoing in terms of the above data for Alaska, Miami, and the U.S. population, the crude death rate for Miami (8.92/1,000) can be expressed as a weighted average of the age-specific death rates (1.19, 0.71, etc. per 1,000) for Miami, where the weights are the population proportion in each age stratum (114,350/562,887, 80,259/562,887, etc.). Similarly, the crude U.S. death rate (8.13/1,000) can be expressed as a weighted average of the U.S. age-specific death rates (1.34, 0.58, etc. per 1,000) with weights consisting of the age distribution in the U.S. population (23,961/91,028, 15,420/91,028, etc.). Therefore, if the U.S. as a whole had experienced the death rates shown above for Alaska, then the crude 1970 U.S. death rate would be 6.71 deaths/thousand, i.e., the directly standardized death rate for Alaska.

[Aside: A technical issue that Rothman and Greenland point out but which we will not worry about is that when the above rates are computed using person-years, rather than people, changes in the death rates can lead to changes in person-years. Unless the death rates are the same across all age strata or the changes in person-years do not change the proportional age distribution, then hypothetical statements such as "if the U.S. as a whole had experienced the death rates shown above for Alaska" require the assumption that replacing the death rates would not alter the proportional age distribution.]

Reasons for standardizing rates

Two main motivations encourage the use of standardized rates. First, summary indices from two or more populations are more easily compared than multiple strata of specific rates. This becomes

especially important when comparing rates from several populations or when each population has a large number of strata. Second, small numbers in some strata may lead to unstable specific rates. When sample populations are so small that their strata contain mostly unstable rates and zeroes, the direct standardization procedure may not be appropriate and an alternate procedure (see below) becomes desirable.

Although standardized rates can summarize trends across strata, a considerable amount of information is lost. For example, mortality differences between two populations may be much greater in older ages, or rates for one population compared to another may be lower in young ages and higher in older ages. In the latter case, a single summary measure obscures valuable information and is probably unwise. Furthermore, different standards could reverse the relative magnitude of the standardized rates depending on which age groups were weighted most heavily. The trade-off between detailed information and useful summarization runs through epidemiologic data analysis methods.

Simultaneous adjustment

Rates can be standardized for two or more variables simultaneously. Table 2 compares age and baseline diastolic blood pressure (DBP)-specific incidences of elevated blood pressure (DBP > 90 mm Hg) in light and heavy subjects (relative weight greater and less than 1.25, respectively) among individuals with DBP previously below 90 mm Hg. The combined population is used as the standard to adjust for age and baseline blood pressure differences in the two weight categories. Computations for simultaneous adjustments are essentially identical to those for the single case:

Standardized rate for low weight subjects

$$= [(0.14 \times 80) + (0.31 \times 59) + ... + (0.11 \times 36)] / 349 = 0.14$$

Standardized rate for heavier subjects

$$= [(0.30 \times 80) + (0.30 \times 59) + ... + (0.59 \times 36)] / 349 = 0.36$$

In this example, the directly standardized rates differ little from the crude rates.

Table 2
Incidence of High Blood Pressure by Baseline Relative
Weight, Blood Pressure, and Age in Evans Co., Ga.

| | Baseline Diastolic Blood Pressure | Relative weight | | | | | | | | |
| | | Light | | | Heavy | | | Total | | |
Age		No.	Cases	Rate	No.	Cases	Rate	No.	Cases	Rate
	Low	70	10	0.14	10	3	0.30	80	13	0.16
25-34	Normal	49	15	0.31	10	3	0.30	59	18	0.31
	Moderate	13	5	0.38	5	4	0.80	18	9	0.50
	Low	67	3	0.04	5	2	0.40	72	5	0.07
35-44	Normal	66	4	0.06	18	4	0.22	84	8	0.10
	Moderate	19	2	0.11	17	10	0.59	36	12	0.33
Total		284	39	0.14	65	26	0.40	349	65	0.19

Spreadsheets

Those of you who are familiar with spreadsheet program (e.g., Lotus 123®, Quattro Pro®, Microsoft Excel®) will readily see the resemblance of the above layout to a spreadsheet. Indeed, spreadsheets are a very convenient method for carrying out a modest number of standardizations. Spreadsheet neophytes will certainly want to learn this method, and even experienced spreadsheet users (who will no doubt want to try this on their own before reading further) may find that creating an age standardization worksheet helps them to learn and understand standardization methods better.

To create the above table in a spreadsheet program, copy the layout, the columns and rows that contain the labels ("35-44", "Moderate", "Light", etc.) and fill in the cells in the first two columns labeled "No." and the two columns labeled "Cases" — but for simplicity of exposition below, do not set aside rows for blank space or horizontal rules or blank columns as separators. If the age categories are placed in column A and the DBP categories in column B, then columns C, D, F, and G (leaving E for the first "Rate" column and "H" for the second) will contain the data for number of participants and number of cases. I will assume that the first row of data (for ages 25-34 years, low diastolic blood pressure) is row **14** (allowing some blank rows for labels and documentation).

To compute the total columns, insert the formula "=C14+F14" into cell **I14** (this corresponds to the number 80 in the table). Upon completing this operation you should see that number appear. Then copy this formula to the rest of the cells in this column (I15-I19) and in the next one (J14-J19). Now, have the spreadsheet compute the row containing the totals for these columns, by using your spreadsheet's summation function to sum the cells in each column. If no rows have been skipped, the summation function will go into cell **C20** and will look something like "@SUM(C14..C19)" [Lotus 123®] or "=SUM(C14:C19)" [Excel®]. Once again you should see the correct total. Then

copy this cell to the other columns to be totaled (D20, F20, G20, I20, J20). (Note to spreadsheet neophytes: spreadsheet programs generally use "relative addressing" by default, so when you copy the formula the program generally adjusts the row and/or column numbers accordingly. Sometimes that's *not* what you want to happen, but in this case it is.)

Then fill in the columns labeled "Rate" by inserting a formula for the ratio of the appropriate "Cases" cell and "No." cell. A simple method for doing this, though not the most elegant, is the following. If the top row of data (25-34 years old, low) is row **14**, the first "No." column is **C**, and the first "Cases" column is **D**, then insert the formula "=D14/C14" into cell **E14** (no rocket science that!). Copy this formula to the remaining cells in the column (E15-E19), and then copy this column to the other two columns labeled "Rate". Your worksheet should now look the same as the table, and you are ready to compute the directly standardized rates.

There are several equivalent ways to proceed. Try this one and then see if you can figure out some others. In the first "Rate" column (i.e., E), a few lines below the "Total" row (e.g., row 26), type in the formula "=E14*I14" (this goes in cell **E26**). This formula multiples the rate for participants who are younger, have low DBP, and are in the lighter relative weight category (**E14**) by the total number of participants who are age25-34 years and have low DBP (**I14**). Then copy E26 to cells E27-E31 and H16-H31.

Each of the latter cells now shows what we might term the "expected number of cases that would occur in each age-DBP stratum of the total participant group if the total group experienced the incidence rates for the lighter-weight participants [for the values in column E] or for the heavier-weight participants [for the values in column H]". Thus, we have only to sum these expected numbers and divide by the total population size. Copy one of the cells that contains a summation function (e.g., C20) to the cell (**E32**) just under the first new column and then copy it (from either C20 or E32) to H32. If the relative addressing works properly, the summation functions should become "=SUM(E26:E31)" and "=SUM(H26:H31)" (or their equivalent in your spreadsheet program). Finally, perhaps on the following row, insert the formulas "=E32/I19" in column E (i.e., in cell **E33**)and "=H32/I19" in column H. You should see the directly standardized rates 0.14 and 0.36, respectively.

If you have faithfully followed the above instructions, you will probably think this is a lot of work to go through for a several-minute task with paper, pencil, and a calculator — even if you have not encountered any difficulties (of your own making or mine). However, this spreadsheet can easily be modified to compute standardized rates for other data, so if you can find it when the need arises it may come in very handy. For now, though, it's probably worthwhile using both calculator and spreadsheet in order to master the computations and concepts.

Standardized ratios and differences

Rates that have been standardized by the direct method, using the same standard population, may be compared in relative or absolute terms (i.e., as a ratio or as a difference). For example, we can obtain a "Standardized Rate Ratio" ("SRR") by dividing the (directly) standardized rate for Miami by that of Alaska. Using the values computed above:

$$\text{SRR} = \frac{\text{directly standardized rate for Miami}}{\text{directly standardized rate for Alaska}} = \frac{6.92}{6.71} = 1.03$$

Similarly, the difference of the two rates would be a "standardized rate difference" (SRD = 6.92–6.71=0.21 [per 1,000 – the ratio has no need for the scaling factor, but the difference does). Since the rates are virtually identical, the SRR is close to 1.0, and the SRD is close to zero, all give the same message: the mortality experience in Alaska, Miami, and the total U.S. are all about the same when the differences due to age structure are eliminated.

In addition, a directly standardized rate can be compared to the crude rate in the population from which the weights were taken (the "standard population"). The reason that this works is that, as noted above, the crude rate for a population can be expressed as a weighted average of the population's stratum-specific death rates (R_k) weighted by its own age distribution. Therefore the crude rate and the directly standardized rates are all weighted averages based on the same set of weights (the proportional age distribution in the standard population). So the following SRR is legitimate:

$$\text{SRR} = \frac{\text{directly standardized rate for Alaska}}{\text{directly standardized rate for total U.S.}} = \frac{6.92}{8.13} = 0.852$$

Standardized ratios and differences are also weighted averages [optional]

It may or may not be of interest to know that the standardized ratios and differences obtained by taking the ratios and differences of directly-standardized rates are also weighted averages. For example, the SRR can be written as:

$$\text{SRR} = \frac{\sum (r_k W_k)}{\sum_j (r'_j W_j)} = \frac{\sum (r_k/r'_k)(r'_k W_k)}{\sum_j (r'_j W_j)} = \frac{\sum [(RR_k)(r'_k W_k)]}{\sum_j (r'_j W_j)}$$

$$= \sum_k \left[(RR_k) \left(\frac{(r'_k W_k)}{\sum_j (r'_j W_j)} \right) \right] = \sum_K (RR_k W'_k)$$

where the RR_k are the stratum-specific rate ratios and the expression in parenthesis is the stratum-specific weight, W'_k for the SRR.

Nonuniformity of stratum-specific rates

Before computing and reporting standardized measures, we should ask the question that applies to any summary measure: does the summary conceal important heterogeneity. If one population has higher rates in some strata but lower rates in others, and stratum sizes are large enough for these differences to be worth paying attention to, then a comparison of standardized rates for the two populations could conceal an important feature of the data. In such a situation, it is important to report the nonuniformity of the stratum-specific rate comparisons and to consider whether computing standardized rates and ratios serves any purpose.

Sparse data

Even though standardized rates can be computed, they are not always meaningful. Use of the same set of weights to average the stratum-specific rates guarantees comparability, but for the comparisons to be meaningful there must also be large enough numbers in all important strata ("important" means those constituting substantial weight in the standardization procedure). Otherwise the stratum-specific rate estimates will be too unstable (i.e., imprecise), and weighting them may only amplify that instability. For example, a rate of 0.10 based on two cases becomes only half as large, 0.05, if two more cases are found. Although the difference between these two rates is small, if they happened to fall in a stratum for which the standard population had a particularly large proportion, then this small difference would be magnified (relative to the other rates) in the standardized rate. There are various rules of thumb for what constitutes "large enough", such as at least 10 or 20 events (e.g., deaths, cases) and a denominator of at least 100, though a specific situation might call for substantially larger numbers.

Indirect standardization

When stratum-specific numbers are small, as is often the case in such populations as a single industrial plant or a small city, stratum-specific rate estimates are too susceptible to being heavily influenced by random variability for the direct standardization method to be satisfactory. Instead, an "indirect" standardization procedure is often used and a "standardized mortality ratio" ("SMR") computed. (The standard mortality difference, computed as the indirectly standardized rate minus the crude rate from the standard population, is also theoretically of interest).

Indirect standardization avoids the problem of imprecise estimates of stratum-specific rates in a study population by taking stratum-specific rates from a standard population of sufficient size and relevance. These rates are then averaged using as weights the stratum sizes of the study population. Thus, the procedure is the mirror-image of direct standardization. In *direct standardization*, the study population provides the rates and the standard population provides the weights. In *indirect standardization*, the standard population provides the rates and the study population provides the weights. (For this reason Ollie Miettinen employs the terms "externally standardized" and "internally standardized", respectively, for what we are calling direct standardization and indirect standardization.)

	Study population	Standard population
Directly-standardized rate	Rates	**Weights**
Indirectly-standardized rate	**Weights**	Rates

We have seen that directly-standardized rates (computed using the same standard population) can be readily compared to each other and to the standard population, because all are based on the set of same weights (those from the *standard population*). **However**, comparison of indirectly-standardized rates can be problematic, because each study population's standardized rate is based on its own set of weights. In fact, the only comparison that is always permissible is the comparison between the study population and the standard population, since these indirect rates are both based on weights from the study population.

Directly-standardized rates are based on one set of weights;
indirectly-standardized rates are based on multiple sets of weights

	Study pop. A	Study pop. B	Standard population
Directly-standardized rate	Rates-A	Rates-B	**Weights**
Indirectly-standardized rate	**Weights-A**	**Weights-B**	Rates

As the above table illustrates, the directly-standardized rates for the three populations are based on the same set of weights (the age distribution of the standard population), but the indirectly-standardized rate for each study population is based on its own age distribution. The resulting lack of comparability of indirectly standardized rates (and of SMR's) is often overlooked or ignored, and as long as the study populations have similar age distributions then there is not necessarily a practical problem. However, if the age distributions differ importantly across the study populations, then comparison of the indirectly-standardized rates could be no better than comparison of the crude rates themselves. Of course, all of these points hold for standardization by other variables; age is used here simply as an example.

Carrying out indirect standardization

Indirect standardization can be thought of as taking the observed number of deaths or events in the study population and comparing that number to an "expected" number of deaths, i.e., the number of deaths that would be expected in the study population if its mortality experience (its stratum-specific rates) were the same as for the standard population. The ratio of observed to expected deaths is termed the Standardized Mortality Ratio (or Standardized Morbidity Ratio if disease, rather than death, is the outcome), abbreviated SMR, and it, rather than standardized rates, is the usual product of the indirect standardization procedure.

The **expected number of deaths** is obtained as follows:

$$\begin{matrix} \text{Expected} \\ \text{number} \\ \text{of deaths} \end{matrix} = \sum \left(\begin{matrix} \text{[Stratum-specific rates from} \\ \text{the standard population]} \end{matrix} \times \begin{matrix} \text{[stratum sizes from} \\ \text{the study population]} \end{matrix} \right)$$

$$= \sum (R_k n_k)$$

and the observed number of deaths is $\sum d_k$

so that

$$\text{SMR} = \frac{\text{Observed deaths}}{\text{Expected deaths}} = \frac{\sum d_k}{\sum (R_k n_k)}$$

where d_k = number of deaths in the k-th stratum of the *study* population ("observed deaths")

n_k = size of the k-th stratum of the *study* population

R_k = death rate in the k-th stratum of the *standard* population

The number of observed deaths can also be expressed as the sum of stratum-specific death rates multiplied by stratum sizes:

$$\begin{matrix} \textbf{Observed} \\ \textbf{number} \\ \textbf{of deaths} \end{matrix} = \sum \left(\begin{matrix} \text{[Stratum-specific rates from} \\ \text{the } \textit{study} \text{ population]} \end{matrix} \times \begin{matrix} \text{[stratum sizes from} \\ \text{the } \textit{study} \text{ population]} \end{matrix} \right)$$

$$= \sum (r_k n_k)$$

where: r_k = death rate in the k-th stratum,

Thus, the SMR can be readily expressed as a ratio of two weighted averages of stratum-specific death rates, where the weights are the proportionate stratum sizes of the *study* population:

$$\text{SMR} = \frac{\text{Observed deaths}}{\text{Expected deaths}} = \frac{\sum (r_k n_k)}{\sum (R_k n_k)} = \frac{\sum (r_k w_k)}{\sum (R_k w_k)}$$

where n_t is the total size of the study population and w_k gives the proportionate stratum sizes, computed as n_k/n.

The SMR indicates the relative excess or decrement in the actual mortality experience in the study population with respect to what might have been expect had it experienced the force of mortality in the standard (or reference) population. [The denominator of the SMR is not precisely the "expected

mortality" when the stratum sizes are in person-years (see Rothman and Greenland, 1998:234, but for our purposes it is close enough.]

Comparison of SMR's

As noted above, the comparison of SMR's (or, equivalently, indirectly-standardized rates) from different study populations is complicated by the fact that the weights used in obtaining the indirectly standardized rates are the stratum sizes of the individual study populations rather than of a (common) standard population. Technically, therefore, one cannot compare SMR's unless the distribution of the standardization variable (e.g., age) is identical across the study populations, in which case standardization is unnecessary since the crude death rates could have been compared directly. Even if two populations have identical stratum-specific rates and therefore their *directly* standardized rates are identical, their *indirectly* standardized rates can be quite different (see example below). Remember, however, that the usual reason for using indirect standardization is that the stratum-specific rate estimates are very imprecise, making directly standardized rates problematic.

Strictly speaking, SMR's can be validly compared across populations with different age distributions in only *one special case*—the situation where the stratum-specific rates in each population are uniform, i.e., they do not vary by age. In this case the weights or age distribution is irrelevant: the average of a set of identical rates will always be the same regardless of the set of weights that are used. If the stratum-specific rates or ratios are reasonably uniform—and if they are widely disparate the usefulness of a single average is somewhat questionable—then a comparison of indirectly standardized rates may be reasonable though admittedly technically improper. If the rates are uniform, however, then the weighting will make little difference so there may be no need to standardize at all.

The following example provides a numerical illustration of the problem of comparing SMR's:

Table 3
Death rates by age in two occupations and a standard population

Age	Occupation A			Occupation B			Standard population		
	Persons	Deaths	Rate	Persons	Deaths	Rate	Persons	Deaths	Rate
40-49	1,000	2	0.002	5,000	10	0.002	30,000	30	0.001
50-59	5,000	20	0.004	1,000	4	0.004	40,000	120	0.003
Total	6,000	22		6,000	14		70,000	150	

$$\text{SMR} \quad \left| \quad \frac{22}{(0.001)(1,000)+((0.003)(5,000)} \quad \right| \quad \frac{14}{(0.001)(5,000)+(0.003)(1,000)}$$

$$\qquad\qquad\qquad 1.38 \qquad\qquad\qquad\qquad\qquad 1.75$$

Though both occupations have exactly the same stratum-specific rates, their SMR's differ, due to the substantially different age distributions for the two occupations. However, the directly standardized rates for both occupations are, reassuringly, the same:

Directly standardized rate for A = (0.002 x 30,000 + 0.004 x 40,000) / 70,000 = 0.0031

Directly standardized rate for B = (0.002 x 30,000 + 0.004 x 40,000) / 70,000 = 0.0031

Similarly, the SRR for each occupation relative to the standard population is 0.0031/0.0021 = 1.48, indicating a 48% higher age-standardized rate of death in each occupational population compared to the standard population. However, the apparent equivalence of the directly standardized rates is misleading. With so few deaths in the younger age stratum in Occupation A and in the older age stratum in Occupation B, the rate estimates are very unstable. In other words, we cannot really estimate some of the rates, so direct standardization is a dubious procedure. Given the substantial uncertainty about what the stratum-specific rates really are, the only conclusion we can be confident of is that both occupations have elevated mortality rates compared to the standard, or reference population. Without assumptions or additional information, we have no evidence from standardization to conclude that one of the occupations is more hazardous (or is not more hazardous) than the other.

Indirectly Standardized Rates (optional topic)

Though not commonly seen, an indirectly standardized rate can be obtained from an SMR as follows:

$$\text{Indirectly-standardized rate} \quad = \text{SMR} \ \times \ \left(\begin{array}{c} \text{Crude death rate in} \\ \text{the } \textit{standard} \\ \text{Population} \end{array} \right)$$

The logic for this relationship is that the SMR gives a standardized comparison of the mortality experience in a study population compared to that in the standard population. So, for example, if the study population has twice the mortality rate of the standard population, the standardized rate for the study population should be twice the observed (crude) death rate in the standard population.

An alternate (and algebraically equivalent) strategy is to multiply the crude death rate from the study population by a "standardizing factor" consisting of the ratio of the crude rate in the standard population to an "index death rate". This "index death rate" is the death rate that would be expected in the study (index) population, due to its age distribution, if in each stratum the corresponding death rate from the standard population applied, i.e., the expected number of deaths divided by the study population size.

$$\text{Indirectly-standardized rate} \quad = \quad \begin{array}{c} \text{Crude death rate} \\ \text{in the } \textit{study} \\ \text{Population} \end{array} \quad \times \quad \text{Standardizing factor}$$

$$= \quad \begin{array}{c} \text{Crude death rate} \\ \text{in the } \textit{study} \\ \text{population} \end{array} \quad \times \quad \frac{\text{Crude death rate in the standard population}}{\text{Index death rate}}$$

where the index death rate is:

$$= \sum \left(\begin{array}{c} \text{Stratum-specific} \\ \text{rates in the} \\ \textit{standard} \text{ population} \end{array} \times \frac{\text{Stratum sizes from study population}}{\text{Total size of study population}} \right)$$

Algebraically, this may be written:

$$\text{Indirectly-standardized rate} \quad = \quad r \times \frac{R}{\sum (R_k n_k)/n} \left[\frac{\text{Crude death rate in the standard population}}{\text{Index death rate}} \right]$$

and may be reformulated:

$$\text{Indirectly-standardized rate} \quad = \quad \frac{d}{\sum (R_k n_k)} \times R$$

$$\text{Indirectly-standardized rate} \quad = \quad \text{SMR} \times R$$

where:

R = crude death rate in the standard population

R_{k_i} = death rate in the k-th stratum of the standard population

r = crude death rate in study population

n_k = size of the k-th stratum of the study population

n = size of the study population

d = total deaths in the study population

Example:

_____

If we use the U.S. rates (from table 1) as a standard, the indirectly standardized death rate for Miami is:

$$\text{Indirectly standardized rate} = \frac{5{,}022}{(1.34^* \times 114{,}350) + (0.58^* \times 80{,}259) + \ldots + (49.51^* \times 92{,}168)} \times 8.13$$

$$= 6.84 \text{ deaths/thousand}$$

*(Per 1,000 population)

For Alaska, the indirect standardized rate is:

$$\text{Indirectly standardized rate} = \frac{285}{(1.34^* \times 37{,}164) + (0.58^* \times 20{,}036) + \ldots + (49.51^* \times 2{,}077)} \times 8.13$$

$$= 7.32 \text{ deaths/thousand}$$

*(Per 1,000 population)

The indirectly standardized rate can be viewed as the study population's crude death rate standardized for the relative "*a priori* mortality proneness" of the study population versus the standard population.

(Returning to basics here)

Table 4
Crude and Age-Standardized* 1970 Death Rates Per 1000 for White Females in Alaska, Miami, and the U.S.

	Alaska	Miami	U.S.
Crude	2.67	8.92	8.13
Direct	6.71	6.92	-
Indirect	7.23	6.84	-

*Standard population is 1970 U.S. white females

Table 4 summarizes the results and indicates that the type of standardization makes a modest difference in this example; the directly standardized rates for Miami and Alaska are closer than their indirect counterparts.

Notice that the age-specific rates from Alaska and Miami do not enter the indirect standardization computations at all. The information which they contain enters indirectly (hence the procedure name), since the observed number of deaths is partly determined by the age-specific rates. But the observed number of deaths is also determined by the stratum sizes.

Choice of Standard Population

Standardized measures describe a hypothetical state of affairs, which is a function of the standard population chosen. For direct age-standardization, the total U.S. population from the previous census is especially common. Since rates standardized to the same external standard are comparable, the selection of a commonly used standard has advantages when comparing rates across different studies. Sometimes investigators compute directly standardized rates based upon one of their own study populations as the standard or by combining two or more study populations to create a standard. But rates standardized to a specific study population are not as readily compared to rates from other studies.

When a study involves a comparison with a "control" population, the choice of a standard should reflect the study goals. For example, an examination of county mortality variation within a state might compare county mortality to the state as a whole. A clean industry may be a good standard for an industrial population exposed to suspected occupational health hazards. Since indirectly standardized measures require knowledge of stratum-specific rates in the standard, data availability constrains the choice.

The choice of a standard population is not always obvious, and there may not be a "best" choice. For example, in comparing syphilis rates across counties in North Carolina, Thomas et al. (1995) decided to standardize the rates by age and sex to reduce the influence of different age-sex distributions in different counties. One obvious choice for a set of weights was the age-sex distribution of North Carolina as a whole. However, another possible choice was to use the age-sex distribution for the U.S. as a whole, so that other investigators could more readily compare syphilis rates in their states to the rates presented in the article. Was there a "right" answer? In this case the choice between the two standards could be regarded as a choice between greater "relevance" and broader comparability. The net result makes little difference, however, since the age-sex distribution of North Carolina and the entire U.S. are very similar. In other situations, however, the choice of standards can indeed change the message conveyed by the results.

Just as the growth of knowledge leads to revisions to disease classification systems, thereby complicating comparisons across revisions, changes in the age distribution over decades creates the dilemma of switch to a new standard population to reflect the present reality versus retaining the existing standard to preserve comparability across time. For this reason mortality rates in the United States have been standardized to the 1940 population distribution almost to the end of the 20th century. Other standards (1970, 1980) were also in use, however, complicating comparisons of mortality statistics. During the 1990's, the U.S. National Center for Health Statistics (NCHS/CDC) coordinated an effort among federal and state agencies to adopt the year 2000 projected U.S. population for standardization of mortality statistics. In August 1998 all U.S. Department of Health and Human Services (DHHS) agencies were directed to use the 2000 Standard Population for age adjusting mortality rates beginning no later than data year 1999 (Schoenborn et al., 2000).

Since the age distribution in 2000 is shifted to the right (older ages) compared to the 1940 population, mortality rates standardized to the 2000 population will be higher than if they were standardized to the 1940 census because they will assign more weight to older age strata, where mortality rates are high. In the same way, comparisons (e.g., ratios) of standardized rates will reflect

the situation among older age groups more than in the past. To be sure, the switch will make comparisons to past data problematic, though NCHS will recompute age-standardized mortality rates for past years based on the 2000 population standard.

The opposite result will occur when at some point it is decided that in a global society all countries should standardized their rates to the World population, to facilitate comparison across countries. Since the large majority of the world's population live in developing countries and is much younger than the population of the U.S. and other developed countries, standardization using a world standard will yield lower standardized rates for most causes of death. As illustrated by the fruit stand example in the beginning of this chapter, different standards can give different, but correct, results. Comparisons, the usual goal of examining rates, may be less affected then the rates themselves, as long as the patterns (e.g., rise in mortality rate with age) are the same in the populations being compared. When that is not the case, then the question of whether it is meaningful to compare summary measures at all becomes more important than the question of which weights to use.

Key concepts

- Populations are heterogeneous – they contain disparate subgroups. So any overall measure is a summary of values for constituent subgroups. The underlying reality is the set of rates for (ideally homogenous) subgroups.

- The observed ("crude") rate is in fact a weighted average of subgroup-"specific" rates, weighted by the size of the subgroups.

- Comparability of weighted averages depends on similarity of weights.

- "Standardized" (and other kinds of adjusted) measures are also weighted averages, with weights chosen to improve comparability.

- Crude rates are "real", standardized rates are hypothetical.

- The "direct" method (weights taken from an external standard population) gives greater comparability but requires more data.

- The "indirect" method (weights taken from the internal study population) requires fewer data but provides less comparability.

- Choice of weights can affect both rates, comparisons of rates, and comparability to other populations, so the implications of using different possible standard populations should be considered.

- Any summary conceals information; if there is substantial heterogeneity, the usefulness of a summary is open to question.

Bibliography

Rothman, Kenneth - *Modern Epidemiology*, Chapter 5 and pages 227-2290. Lilienfeld and Lilienfeld - *Foundations of epidemiology*, Measures of mortality, pp. 71-80; Mausner & Kramer - *Epidemiology: an introductory text*, pp. 338-344.

Inskip, Hazel; Valerie Beral, and Patricia Fraser. Methods for age adjustment of rates. *Statistics in Medicine* 1983; 2:455-4660.

Gaffey WR: A critique of the standardized mortality ratio. *J Occup Med* 18:157-160, 1976.

Schoenborn, Charlotte; Richard Klein, Virginia Fried. Age adjustment of National Center for Health Statistics data using the 2000 projected U.S. population with emphasis on survey data systems. Division of Health Interview Statistics, NCHS/CDC. To appear in the *Healthy People Statistical Note* series (http://www.cdc.gov/nchs/products/pubs/workpap/ageadjust.htm, 9/10/00).

Thomas, James C.; Alice L. Kulik, Victor J. Schoenbach. Syphilis in the South: rural rates surpass urban rates in North Carolina. *Am J Public Health* 1995; 85:1119-1122

Tsai SP, Hardy RJ, Wen CP. The standardized mortality ratio and life expectancy. *Am J Epidemiol* 1992; 135:824-831.

Appendix on Standardized Mortality Ratios

(courtesy of Raymond Greenberg, M.D.,Ph.D.

I. DEFINITION. The Standardized Mortality Ratio (SMR) is a measure of mortality in a study population, relative to mortality in a reference population. The SMR answers the following question: "How does the number of observed deaths compare with the expected number of deaths, if our study group had the age-specific mortality rates of the reference population during these study years?"

II. CALCULATION. In principle, any reference population yielding sufficiently precise rates can be used to obtain the expected death number, but it is customary to use the general population. The SMR is given by the following expression:

$$\text{SMR} = \frac{\text{Observed deaths in study population}}{\text{Expected deaths in study population}}$$

The SMR is usually scaled up by multiplying it by 100. An SMR over 100 indicates that more deaths were observed than expected (i.e., the study population had a relatively poor outcome). An SMR less than 100 means that fewer deaths were observed than expected (i.e., the study population had a relatively favorable outcome). Obviously, the value of the SMR will depend on the choice of the reference population used for the comparison mortality rates. If the reference population is healthy, they will have low mortality rates and thereby increase the SMR. Conversely, if the reference population is unhealthy, they will have high mortality rates and thereby decrease the SMR. It is therefore crucial to choose an appropriate reference population or at least to know in which direction the reference population differs from an appropriate one.

III. HEALTHY WORKER EFFECT. The SMR is frequently used to examine mortality in an industrial plant or industry. However, when workers are compared to the general population, it is common to find lower mortality rates in the workers (SMR less than 100). The reason is thought to be that the general population includes people who are too sick to work. The elevated mortality in such people raises the mortality rate of the general population, so that mortality in the general worker population is lower. This phenomenon is called the healthy worker effect. The healthy worker effect is an important consideration primarily for mortality from diseases, such as cardiovascular disease, where an extended period of physical limitation or disability frequently precedes death and thus affects entrance into and remaining in the workforce.

IV. SAMPLE CALCULATION: Suppose you are studying male textile workers between the ages of 20 and 39 years between the years 1960 and 1979.

a. Observed deaths

Age	Period 1960-1969	1970-1979	Total
20-29	1	2	3
30-39	2	3	5
Total			8

b. Person-years of exposure

Age	Period 1960-1969	1970-1979	Total
20-29	1,000	500	1,500
30-39	500	1,000	1,500

c. Mortality rates from reference population

Age	Period 1960-1969	1970-1979
20-29	1/1,000py	2/1,000py
30-39	2/1,000py	4/1,000py

d. Expected deaths (b x c)

Age	Period 1960-1969	1970-1979	Total
20-29	1	1	2
30-39	1	4	5
Total			7

$$\text{SMR} = \frac{\text{Observed Deaths in Study Population}}{\text{Expected Deaths in Study Population}} \times 100 = \frac{8}{7} \times 100 = 114$$

or a 14% elevation in mortality.

V. CAUTIONS IN USE OF SMR:

a. An SMR is an indirect standardization procedure (standard rates applied to study population) and therefore two SMR's cannot be compared, unless their respective populations have the same age distribution (in which case, why standardize). [If the age distributions are not markedly different or the relationships in mortality rates between the populations are similar

across age strata, then the damage is not great. The latter possibility can only rarely be checked, of course, since SMR's are typically computed in situations where there are too few deaths in each stratum to calculate meaningful stratum-specific rates.]

b. SMR's do not readily translate into life-expectancy (though recent work provides an approximation).

c. As length of follow-up increases, an SMR based on cumulative mortality tends toward 100.

(See Gaffey WR: A critique of the standardized mortality ratio. *J Occup Med* 18:157-160, 1976