An overview of standardization

A key concept in epidemiology is that virtually every population is heterogeneous with regard to characteristics relevant to the occurrence of disease and other health-related events. These characteristics include sociodemographic (e.g., age, gender, education, religion), geographic, genetic, occupational, dietary, medical history, and innumerable other characteristics. Therefore a population can be thought of as a composite of subgroups (ultimately, subgroups of size one, i.e., individuals, but epidemiologic measures break down at that point). These subgroups, often called **strata**, can be defined by which characteristics are important for the phenomenon under study and for which data are available. If, for example, there are five strata defined by age, each will have a certain number of people (n_i) and number of cases (d_i), so that the overall population values are the sums $N=\Sigma n_i$ and $D=\Sigma d_i$, and D/N is the overall proportion of cases or, if the d_i represent deaths during one year, the annual mortality rate.

This overall rate is said to be **crude**, since it does not take explicit account of the composition of the population. The crude rate summarizes the experience in all of the strata that compose the population, their individual ("**specific**") rates being d_1/n_1, d_2/n_2, d_3/n_3, d_4/n_4, d_5/n_5. The crude rate is the simplest and most straightforward summary of the population experience. But it is important to remember the heterogeneity that underlies the crude rate or, for that matter, any summary measure. This issue is particularly relevant when we proceed to interpret comparisons between groups or across time periods, because if the populations we are comparing differ in composition, then at least some of what we observe may be attributable to compositional differences that are extraneous to the question at hand.

As indicated, crude rates are summaries, summaries of the rates for the component subgroups, with each subgroup's rate being represented in proportion to the subgroup's size (a "democratic" approach, if you like). If, as is often the case, we are comparing groups whose subgroups have different relative sizes, then the comparison of their crude rates is confounded by their compositional difference. 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); 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 convenient to have summary measures that could be used for 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, whereas the supermarket average price gives more weight to the price of apples. We're comparing apples to mangoes, instead of fruit stand to supermarket.

Clearly what we need is a procedure that weights the prices in the same way for each seller, so that each average price gives 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 seller will prefer that you assign the major weight to the price of mangoes, so that her enterprise will appear to be 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 lower than the fruit stand's average price (with zero weighting for mangoes).

Which set of weights is correct? People who don't like mangoes might agree with the supermarket owner. People who like mangoes – or fruit stands – would not. As long as the data are not so sparse as to create instability in estimates (as discussed below), the choice of weights (a.k.a. **standard population**) depends upon what summaries are being compared (and whether others may later be compared to these) and on the context of the comparison. It is also advisable to find a rationale for the choice that has arguments in its favor other than that it happens to give you a result that 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.