values of the controls showed a violent contrast between the sample and the whole population. . . .[6]

We will develop this fundamental point further by another example.

EXAMPLE 287 SMOKING AND CANCER

In 1954, widespread publicity was given to the preliminary report of an extensive statistical study by the American Cancer Society of the relation of smoking, especially cigarette smoking, to lung cancer.[7] The study had been made by classifying 187,766 men 50 to 70 years of age according to their smoking habits, and then 20 months later determining which men had died and for what causes. The lung-cancer death rate was about nine times as high for men smoking one pack or more of cigarettes per day as for nonsmokers. A number of earlier but smaller and less publicized studies in several countries showed similar results. The American Cancer Society is still continuing to follow the original group.

A medical statistician reviewing these results noted that the percent of smokers in the sample was less than that in the comparable portion of the U. S. population, that the death rates from all causes for the sample were about 30 percent below the corresponding U. S. rates (even the heavy smokers in the sample showed a lower cancer death rate than the corresponding U. S. rate), and that for the sample, death rates not only from cancer but from all causes were higher for cigarette smokers than for nonsmokers.[8] All these facts indicated to him the presence of substantial selectivity in the sample, reminded him of previous incidents in medical statistics that had proved fiascos, and led him to suggest that a simple selective process like the following *could* have produced a spurious association.

Suppose, first, that the population consists of two groups: Group I, constituting 3 percent of the population, which is on the verge of death and has a death rate during the period of the study of 99 percent; and Group II, constituting 97 percent of the population, which has a death rate of 0.03 percent. Assume that 80 percent of each group are smokers, and that the death rates are the same for smokers and nonsmokers within each group. Thus, the death rates would be (in percents):

6. Jerzy Neyman, *Lectures and Conferences on Mathematical Statistics and Probability* (2d ed.; Washington; Graduate School of the Department of Agriculture, 1952), pp. 105–106.

7. E. Cuyler Hammond and Daniel Horn, "The Relationship between Human Smoking Habits and Death Rates," *Journal of the American Medical Association*, Vol. 155 (1954), pp. 1316–1328.

8. Joseph Berkson, "The Statistical Study of Association between Smoking and Lung Cancer," *Proceedings of the Staff Meetings of the Mayo Clinic*, Vol. 30 (1955), p. 319.

	Group I	Group II
Nonsmokers	99	0.03
Smokers	99	0.03

The relative frequencies in the population are:

	Group I	Group II
Nonsmokers	0.006	0.194
Smokers	0.024	0.776

Using again the formula for weighted means (Sec. 7.4.2), we obtain the following death rates:

Nonsmokers:

$$\frac{0.006 \times 99 + 0.194 \times 0.03}{0.006 + 0.194} = 0.03 \times 99 + 0.97 \times 0.03$$
$$= 3.00 \text{ percent.}$$

Smokers:

$$\frac{0.024 \times 99 + 0.776 \times 0.03}{0.024 + 0.776} = 0.03 \times 99 + 0.97 \times 0.03$$
$$= 3.00 \text{ percent.}$$

Now introduce the assumption that the proportion of smokers cooperating in the study differs between Groups I and II. Suppose that half the individuals in Group I are included in the sample, the omissions being those whose poor health is obvious or prevents their being interviewed, whether they smoke or not. For Group II, assume that 99 percent of nonsmokers and 65 percent of smokers cooperate. Such differential selectivity would result in a death rate in the sample that is 50 percent higher for smokers than for nonsmokers, being 2.33 percent for smokers and 1.55 percent for nonsmokers. This last step is contained in the bottom half of Table 289, which summarizes the entire analysis.

The crucial assumption, which accounts for the results, is that there is differential selectivity between smokers and nonsmokers, and that *the differential is greater for the healthy than the unhealthy.* In such a study many of the very unhealthy are simply not available for interview, and are therefore eliminated on grounds of ill-health, regardless of smoking habits; there is no harm in that alone. The healthy differ in their availability and cooperativeness and it is not at all implausible that more smokers than nonsmokers will refuse to cooperate in a study of the ill-effects of smoking; again, there is no harm in that alone. It is the combined effect that is damaging. In any event, as Berkson says,

... The fact that the exact mechanism of such selective association is not readily visualized is not an adequate reason for considering the suggestion of its possible existence to be—as it has been characterized—"far fetched." ... Nor is it conclusive that the considerable number of statistical studies that have been published all agree in showing an association between smoking and cancer of the lung. On the contrary, ... if correlation is produced by some elements of the statistical procedure itself, it is almost inevitable that the correlation will appear whenever the statistical procedure is used.[9]

TABLE 289

HYPOTHETICAL ILLUSTRATION OF POSSIBLE EFFECTS OF SELECTIVITY IN PRODUCING SPURIOUS ASSOCIATION

Group	Smoker	Number in Group			Number of Deaths			
		Group I	Group II	Both groups	Group I	Group II	Both groups	Rate (percent)
Population	No	600	19,400	20,000	594	6	600	3.0
	Yes	2,400	77,600	80,000	2,376	24	2,400	3.0
	Total	3,000	97,000	100,000	2,970	30	3,000	3.0
Sample	No	300	19,206	19,506	297	6	303	1.55
	Yes	1,200	50,440	51,640	1,188	16	1,204	2.33
	Total	1,500	69,646	71,146	1,485	22	1,507	2.12

Source: Berkson, *op. cit.*, Appendix Table 1.

In bringing up this example, it is not our intention to express or imply an opinion as to whether the association between lung cancer and smoking is spurious. Rather, our intention is to point out the issues involved, for they apply to virtually all associations based on analyses of data from experience rather than experiment.

The example illustrates also the danger of assuming that substantial nonresponse or non-cooperation will have no serious effect on statistical studies. If you try out various assumptions about percentages of differential selectivity on the basic data of Table 289, you will find that the observed association between smoking and lung cancer will be considerably different for different assumed percentages. The rates of selectivity in this example reflect both the original sampling process and the success in obtaining cooperation from those designated by the sampling process. The problem of failure to get information from those designated by the sampling process is called the problem of nonresponse or non-cooperation. The problem of nonresponse may be a serious one even if the original sampling process is completely sound. For example, in the Salk polio vaccine tests of

9. Berkson, *op. cit.*, p. 332. Footnotes omitted.

1954 there is some evidence that children whose parents withheld permission to participate in the trials may have been less susceptible to polio than children who did participate.[10]

Unfortunately, a frequent tendency in practice is to ride roughshod over the nonresponse problem and pretend either that it does not exist or at least that it could not possibly be serious. The most fruitful way of attacking the problem, once it is recognized, is to spend relatively more resources in getting information on the potential nonrespondents by repeated calls, special interviewers, etc. This means that the total number of interviews will be smaller than otherwise, but that the data obtained will be better.

9.5 STANDARDIZED AVERAGES

A method of allowing, or "adjusting," for differences in the composition of groups which are to be compared is to compute what the means would be if the groups had the same standard composition.

In the method of standardized means, the means for the subgroups are combined in the group mean not on the basis of different weights for each group but on the basis of standardized weights. As an illustration, let us turn to Table 277 and for each appearance group combine the three averages of individual years on the basis of equal weights. We have for the beautiful (see Sec. 7.4.2 on computing weighted means)

$$\tfrac{1}{3} \times 1.16 + \tfrac{1}{3} \times 1.57 + \tfrac{1}{3} \times 1.53 = 1.42.$$

Proceeding in the same way for the other appearance groups we find the results shown in the "weights equal" column of Table 291A. This table also shows the results of using as the standard weights the frequencies in the last column of Table 280. With these weights, the standardized mean for the beautiful, for example, is

$$0.339 \times 1.16 + 0.379 \times 1.57 + 0.281 \times 1.53 = 1.42.$$

In this example, unlike some others, there is little difference between the two sets of standardized means, or between the standardized and the unadjusted means. All indicate that after adjusting for the effect of year in college, mean grades are consistently lower when appearance is better. Such standardized means have an advantage

10. K. A. Brownlee, "Statistics of the 1954 Polio Vaccine Trials," *Journal of the American Statistical Association*, Vol. 50 (1955), pp. 1005–1013.