

## **2. RATES AND RATE STANDARDIZATION**

- 2.1 Calculation of age- and calendar period-specific rates
- 2.2 Summarizing a set of rates
- 2.3 Comparative measures of incidence and mortality
- 2.4 Proportional measures of incidence and mortality

## CHAPTER 2

### RATES AND RATE STANDARDIZATION

Analysis of data from a cohort study involves estimation of the rates of cancer and other diseases of interest which occur among cohort members during the study period. Cancer occurrence is most appropriately measured in terms of incidence rates, for example, as the number of newly diagnosed cases per 100 000 person-years of observation time. For a variety of practical reasons, however, most of the important cohort studies discussed in the preceding chapter used death from disease rather than its diagnosis as their principal endpoint. From the point of view of formal statistical analysis, it makes little difference what endpoint is selected, and we refer almost exclusively to 'deaths', 'mortality' and 'survival', leaving it to the reader to make the necessary substitution of terminology (e.g., 'cases', 'morbidity' and 'disease-free survival') as required for incidence data. However, since death is often preceded by a period of ill health, and the health status of the subject may influence his exposure to the agents under investigation, mortality data are subject to particular problems of interpretation, as discussed in §1.5c. Lagging of exposure variables is one means of partially accounting for modification of exposures during the interval between first appearance of disease symptoms and death.

Cancer rates vary widely according to sex, age, calendar time and a number of other demographic variables. We begin the chapter with a description of procedures used to estimate age- and time-specific disease rates from cohort data. The rates may also be specific for sex and race. Methods of estimation of incidence rates using cancer registry material were also discussed in Chapter 2 of Volume 1, which the reader may wish to consult for further elaboration of the concepts of rate and risk and a discussion of alternative methods available for their determination.

Analysis of cohort data typically involves a comparison of the rates observed in the study group with rates for the general population. This is a useful way of identifying diseases which occur at especially high or low frequency in the cohort, so that they may be studied further in relation to particular exposures. Since the age distribution of the cohort will generally be different from that of the population as a whole, and may also be evolving with time, such comparisons are best made on an age-time-specific basis. Thus, the second topic considered in this chapter is the combination of age-time-specific cohort rates so as to facilitate their comparison with standard or general population rates. Direct and indirect standardizations are presented as the two basic methods of summarizing a set of component rates. The corresponding comparative measures, known as the comparative mortality figure (CMF) and the standardized

mortality ratio (SMR), are introduced and discussed in terms of the advantages and disadvantages they offer. We develop methods for evaluating the statistical significance of the observed differences in age-time-specific rates between study and standard populations, and for putting confidence intervals around the comparative measures.

A final section briefly describes some analogous procedures for age adjustment of proportional incidence and mortality measures that are used to evaluate disease frequencies when person-years denominators are not available.

## 2.1 Calculation of age- and calendar period-specific rates

The basic feature of cohort studies that distinguishes them from cross-sectional, case-control or other types of investigation is that, at least in principle, each subject is kept under continuous surveillance for a defined interval of time. If the study endpoint is death, we assume that each subject is 'at risk' of death during the entire interval from his entry into the study until his exit. This means that the study period should contain no interval during which the subject is known to be alive as a condition of cohort membership. If the cohort is defined to consist of all workers with at least five years of employment in a certain factory, therefore, the first five years of their employment history would be excluded from the observation period. A second critical assumption is that any death that actually occurs during the study period will be recorded. For cohorts defined on the basis of past records, this implies that adequate mechanisms exist for tracing individuals from their date of entry into the study until death or until the study's closing date. If no record exists of someone's whereabouts after a certain point in time, he should generally be considered as having left the study at that point. Obvious problems of selection bias exist if such losses are at all frequent, since the causes of and ages at death for 'lost-to-follow-up' subjects may well differ from those for persons who are successfully traced.

The basic method used to estimate age-time-specific mortality rates is to determine for each individual the amount of observation time contributed to a given age  $\times$  calendar period category and to sum up those contributions for all cohort members so as to obtain the total number of person-years of observation in that category. These person-years form the denominators of rates the numerators of which are simply the numbers of deaths due to a given disease, likewise classified by age and calendar year of death. The process is illustrated in Figure 2.1, which shows schematically the course of one worker who was entered on study (point A) at age 43.71 in year 1956.03 and left 11.12 years later (F). He contributed observation time to five separate cells, boundary crossings being made at points B through E. The duration of time spent in each cell is easily determined, as shown in Table 2.1. In some applications, particularly when the observation period is relatively short, the calendar-year axis is ignored and the rates are determined by age interval alone. Computer programs for performing the calculations have been developed by Hill (1972), Monson (1974), Waxweiler *et al.* (1983), Gilbert and Buchanan (1984) and Coleman *et al.* (1986), among others.

Sometimes the exact dates of birth and of entry and exit from study, which are needed to draw Figure 2.1, will not be available. Then, approximate numbers of person-years may be calculated as shown in the right-hand column of Table 2.1, using

Fig. 2.1 Schema showing the follow-up of one person in a cohort study

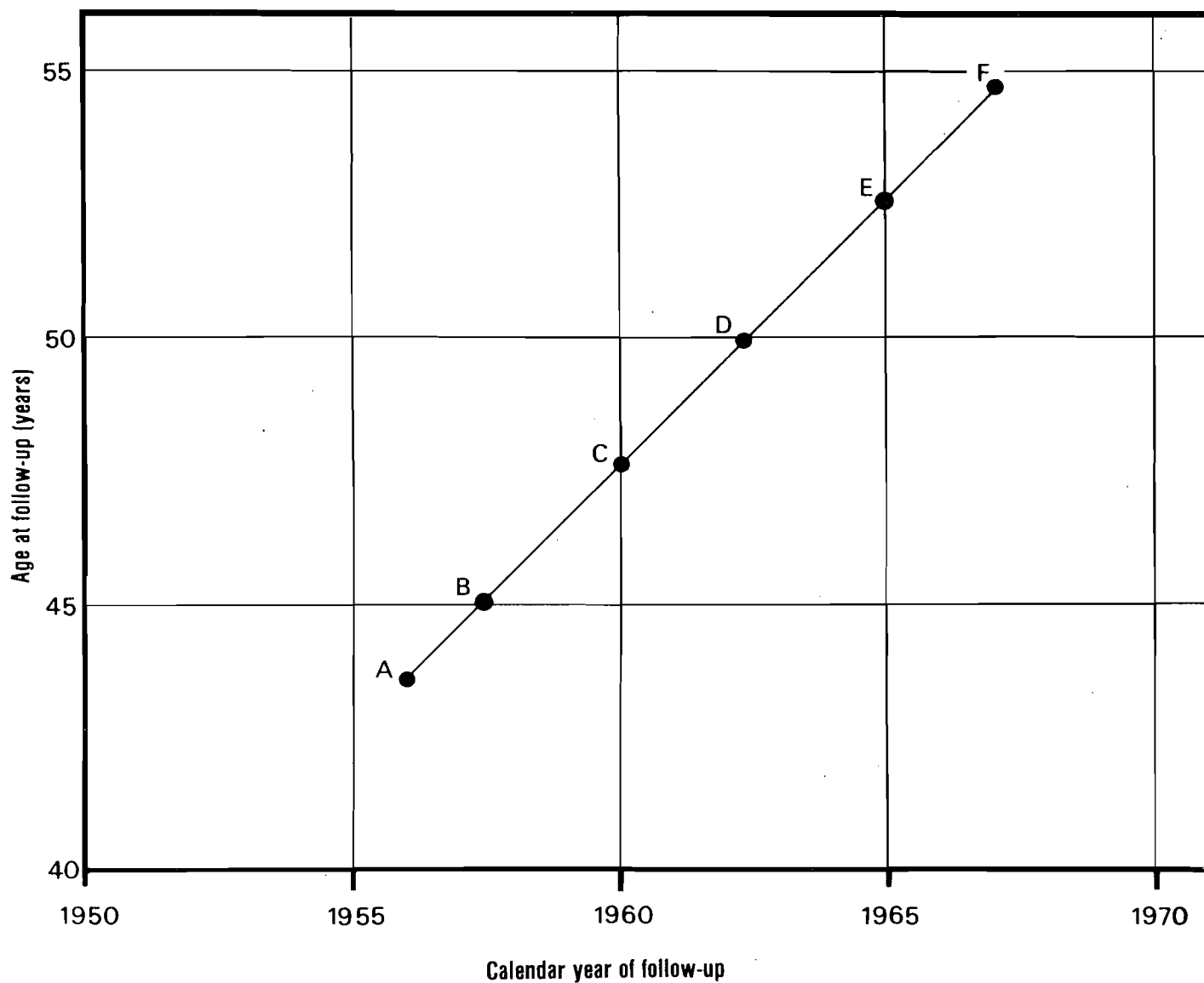


Table 2.1 Calculation of exact and approximate age- and year-specific person-years at risk

Point <sup>a</sup>	Coordinates (year, age)	Quinquennium		Person-years	
		Year	Age	Exact	Approximate
A	(1956.03, 43.71)	1955–1959	40–44	1.29	1.50
B	(1957.32, 45.00)		45–49	2.68	2.00
C	(1960.00, 47.68)	1960–1964	45–49	2.32	3.00
D	(1962.32, 50.00)	1960–1964	50–54	2.68	2.00
E	(1965.00, 52.68)	1965–1969	50–54	2.15	2.50
F	(1967.15, 54.83)				
Total				11.12	11.00

<sup>a</sup> See Figure 2.1

the three integer variables, age at entry, year of entry and year of exit. The approximation is based on the notion that a person aged 43 in 1956 will be 44 in 1957, 45 in 1958 and 54 in 1967. He contributes 0.5 years of observation time to the calendar year of entry (1956), 0.5 years to the year of exit (1967), and a full 1.0 year to each intervening year. There would be a single 0.25-year contribution for someone who enters and leaves the study in the same calendar year. The discrepancies between the exact and approximate figures tend to be averaged out when cumulated over individuals, so that the approximate method is sufficiently accurate for most practical purposes.

Cause-specific national death rates are typically published by five-year intervals of age and calendar year (Case & Pearson, 1957; Grove & Hetzel, 1968). Such 'quinquennia' are widely used in cancer epidemiology, and our example of the calculation of age- and calendar period-specific rates illustrates this standard breakdown. Analogous methods may be used if the age/time intervals are longer or shorter than five years.

### Example 2.1

Appendix IE describes in some detail the design and execution of the Lee and Fraumeni (1969) study of Montana copper smelter workers, in which 8047 male subjects were entered into study on 1 January 1938 if they had worked for at least one year and were still employed on that date, or at the end of their first year of employment for those hired later. Table 2.2 shows the number of person-years in each quinquennium as determined by both exact and approximate methods for 8014 workers on whom full data were available. (Records had been lost for 33 of the original cohort.) These data include the follow-up through 1977 for workers who were still alive and under observation on 31 December 1963, the closing date of the original study (Lubin *et al.*, 1981).

The approximate method of calculation, based on integral ages and years, was modified to account for the fact that the 2517 men who entered the study at the beginning of 1938 were eligible for a full year's observation, whereas those who entered the study later were, on average, observed for only half of the first year. Likewise, nine months of observation during 1977 was counted for those still alive and being followed at the study's end (30 September 1977). Except for a few discrepancies along the boundaries of the table, this adjustment assures that the agreement between exact and approximate calculations is quite good.

Table 2.3 presents the numbers and rates of deaths from all causes classified by age and calendar period. The rates are based on sufficiently large numbers for most cells that they display a reasonable degree of numerical stability. For many specific causes of death, however, the numbers are smaller and a display of the age-period-specific rates in such a detailed manner is not helpful.

## 2.2 Summarizing a set of rates

Large tables of rates confront the investigator with a bewildering array of detail that is difficult to assimilate and utilize effectively. Even if attention is focused on a particular calendar period or column in Table 2.3, consideration of the rates in five-year age categories requires one to look at up to 18 separate numbers. There is a clear need for one or two summary measures that are easy to interpret yet retain most of the essential information in the age-specific data.

This section describes the calculation and interpretation of a commonly used summary measure, the directly standardized rate. We mention at the outset, however, that important information may be lost through use of this and other traditional approaches to data analysis. The remainder of the monograph emphasizes alternative

Table 2.2 Exact and approximate<sup>a</sup> person-years of observation in the Montana cohort, by age and calendar year

Age range (years)	Calendar period									Totals
	1938– 1939	1940– 1944	1945– 1949	1950– 1954	1955– 1959	1960– 1964	1965– 1969	1970– 1974	1975– 1977	
10–14	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2
	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5
15–19	42.6	208.1	387.7	173.8	59.8	0.0	0.0	0.0	0.0	871.9
	68.0	164.0	351.0	123.8	38.0	0.0	0.0	0.0	0.0	744.8
20–24	645.4	2157.6	2001.0	2361.4	1380.5	142.7	0.0	0.0	0.0	8 688.6
	679.0	2301.8	2012.0	2352.0	1349.5	118.5	0.0	0.0	0.0	8 812.8
25–29	690.3	2587.0	3143.6	3359.7	3038.4	1377.1	135.9	0.0	0.0	14 332.0
	682.5	2647.5	3290.8	3361.3	3039.5	1362.5	112.5	0.0	0.0	14 496.5
30–34	689.4	2378.9	3451.4	3941.6	3845.6	2950.1	1342.0	130.8	0.0	18 729.9
	693.0	2354.3	3508.0	4091.0	3844.8	2951.0	1327.5	108.0	0.0	18 877.5
35–39	607.9	2196.9	3091.0	3906.2	4313.5	3718.1	2883.8	1321.5	129.8	22 168.6
	598.5	2186.5	3076.8	3950.5	4456.5	3721.0	2885.0	1309.0	108.0	22 291.8
40–44	482.1	1829.2	2813.8	3603.3	4264.7	4164.2	3632.2	2815.4	941.9	24 546.9
	471.0	1784.5	2826.5	3573.0	4321.3	4296.5	3631.5	2815.5	945.5	24 665.3
45–49	451.4	1431.8	2298.9	3222.4	3851.7	4044.2	3967.8	3501.7	1812.1	24 581.9
	450.0	1426.5	2258.8	3227.8	3810.0	4108.0	4083.0	3505.5	1812.0	24 681.5
50–54	470.0	1230.1	1758.3	2598.0	3432.1	3613.6	3753.0	3728.5	2010.1	22 593.8
	465.0	1226.0	1747.0	2544.3	3437.0	3575.0	3810.0	3840.5	2018.0	22 663.3
55–59	424.5	1209.1	1389.6	1750.3	2560.2	3067.1	3263.4	3433.4	2101.5	19 199.0
	408.5	1206.0	1377.5	1737.5	2505.0	3062.5	3219.5	3486.0	2168.0	19 170.5
60–64	308.4	889.9	1221.2	1283.1	1574.6	2148.7	2535.5	2805.5	1790.9	14 557.9
	315.0	855.8	1217.0	1276.0	1561.0	2097.0	2527.0	2764.5	1799.0	14 412.3
65–69	248.4	667.2	762.9	1031.5	1054.8	1283.1	1699.4	2065.9	1338.7	10 152.1
	231.0	659.0	735.0	1014.0	1042.0	1270.5	1653.0	2056.5	1313.0	9 974.0
70–74	147.2	425.6	496.5	559.1	766.3	796.7	961.2	1325.3	951.2	6 429.1
	137.5	399.0	478.0	529.0	744.5	789.5	952.0	1285.5	941.5	6 256.5
75–79	49.2	215.3	259.5	287.7	324.6	493.6	526.8	647.3	499.9	3 303.8
	45.5	194.0	246.0	274.0	306.5	468.0	515.5	636.0	476.5	3 162.0
80–84	14.6	69.0	125.3	107.3	147.7	148.6	260.6	265.4	191.7	1 330.1
	11.0	63.5	111.0	97.5	140.0	132.5	245.5	253.5	187.0	1 241.5
85+	1.4	13.5	35.5	48.0	30.2	52.7	62.9	135.1	90.3	469.6
	1.0	10.0	31.5	41.0	26.5	47.5	53.5	127.0	81.5	419.5
Totals	5 272.7	17 510.5	23 236.3	28 233.4	30 644.7	28 000.3	25 024.5	22 175.7	11 858.2	191 956.3
	5 256.5	17 479.8	23 266.8	28 192.5	30 622.0	28 000.0	25 016.0	22 187.5	11 850.0	191 871.0

<sup>a</sup> Exact entries listed above approximate ones for each cell

methods of analysis that we believe are preferable for analytical epidemiology, namely the fitting of statistical models to the age- and period-specific rates in such a way that their essential structure is highlighted and purely 'random' fluctuations are identified as such.

(a) *The directly standardized rate*

Direct standardization appears to have been motivated originally by the idea of determining the crude disease rate that would be observed in the cohort if its age distribution were the same as that of the standard population. The directly standard-

Table 2.3 Number of deaths and death rates (per 1000 person-years)<sup>a</sup> from all causes in the Montana cohort, by age and calendar year

Age range (years)	Calendar period									Totals
	1938– 1939	1940– 1944	1945– 1949	1950– 1954	1955– 1959	1960– 1964	1965– 1969	1970– 1974	1975– 1977	
15–19	0	2	0	0	0	0	0	0	0	2
	0.0	9.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.3
20–24	1	6	2	3	4	1	0	0	0	17
	1.5	2.8	1.0	1.3	2.9	7.0	0.0	0.0	0.0	2.0
25–29	2	11	2	4	7	2	1	0	0	29
	2.9	4.3	0.6	1.2	2.3	1.5	7.4	0.0	0.0	2.0
30–34	2	8	6	16	12	10	5	0	0	59
	2.9	3.4	1.7	4.1	3.1	3.4	3.7	0.0	0.0	3.2
35–39	0	10	14	23	16	14	7	2	0	86
	0.0	4.6	4.5	5.9	3.7	3.8	2.4	1.5	0.0	3.9
40–44	1	12	15	27	25	40	11	21	4	156
	2.1	6.6	5.3	7.5	5.9	9.6	3.0	7.5	4.2	6.4
45–49	1	13	21	26	40	40	43	28	15	227
	2.2	9.1	9.1	8.1	10.4	9.9	10.8	8.0	8.3	9.2
50–54	5	25	25	42	61	58	55	52	28	351
	10.6	20.3	14.2	16.2	17.8	16.1	14.7	13.9	13.9	15.5
55–59	1	28	26	48	78	86	82	76	37	462
	2.4	23.2	18.7	27.4	30.5	28.0	25.1	22.1	17.6	24.1
60–64	2	35	35	39	54	83	87	103	50	488
	6.5	39.3	28.7	30.4	34.3	38.6	34.3	36.7	27.0	33.5
65–69	11	29	35	51	56	68	71	83	53	457
	44.3	43.5	45.9	49.4	53.1	53.0	41.8	40.2	39.6	45.0
70–74	5	32	42	43	61	57	56	75	61	432
	34.0	75.2	84.6	76.9	79.6	71.5	58.3	56.6	64.1	67.2
75–79	3	16	30	34	37	55	57	63	46	341
	61.0	74.3	115.6	118.2	114.0	111.4	108.2	97.3	92.0	103.2
80–84	0	7	18	24	26	25	29	39	29	197
	0.0	101.4	143.6	223.7	176.1	168.3	111.3	147.0	151.3	148.1
85+	0	2	11	6	12	11	15	27	16	100
	0.0	147.7	309.5	125.0	397.7	208.7	238.7	199.8	177.2	213.0
Totals	34	236	282	386	489	550	519	569	339	3404
	6.4	13.5	12.1	13.7	16.0	19.6	20.7	25.7	28.6	17.7
Standardized rates (1950 US population aged 40–79 years):										
	12.8	26.5	26.0	27.9	29.5	29.4	25.0	24.4	22.4	
<sup>a</sup> Numbers of deaths are listed above the corresponding rate										

ized rate is obtained by applying the age-specific cohort rates to the standard age distribution. More formally, denote by  $d_j$  the number of deaths in the  $j$ th of  $J$  age groups, by  $n_j$  the person-years denominator, and by  $\hat{\lambda}_j = d_j/n_j$  the corresponding rate. In statistical parlance  $\hat{\lambda}_j$  is known as an estimate (hence the  $\hat{\cdot}$ ) of the 'true' but unknown rate  $\lambda_j$  that would be observed if an infinite amount of data were available. If, in addition,  $w_j$  denotes the number (or proportion) of individuals in the standard

population in the  $j$ th age group, the directly standardized rate is written

$$\hat{\Lambda} = \sum_{j=1}^J w_j \hat{\lambda}_j \quad (2.1)$$

as a weighted sum (or average) of the age-specific rates.  $\Lambda$  denotes the corresponding 'true' quantity.

Table 2.4 shows several idealized populations used for direct standardization of cancer incidence rates (Waterhouse *et al.*, 1976). Since the weights sum to 100 000, the corresponding directly standardized rates calculated from (2.1) will have units of cases per 100 000 person-years. The African population is considerably younger than the European. The world population, which occupies an intermediate position, has long been used by Segi (1960) and associates to standardize cancer mortality data collected by the World Health Organization.

Table 2.5 shows the actual age distributions for 1 000 000 persons in the USA for the years 1950 and 1970. Note the effect of the post-war 'baby boom' on the two age structures. These figures are often used to standardize the mortality rates of US cohorts.

In order to promote comparability between series, we recommend that a published set of weights such as those shown in Table 2.4 or 2.5 be used for direct standardization, rather than an ad-hoc set constructed by the investigator. When world

Table 2.4 Standard populations used for the computation of age-standardized and truncated standardized incidence rates<sup>a</sup>

Age range (years)	African	World	European	Truncated
0–	2 000	2 400	1 600	—
1–4	8 000	9 600	6 400	—
5–9	10 000	10 000	7 000	—
10–14	10 000	9 000	7 000	—
15–19	10 000	9 000	7 000	—
20–24	10 000	8 000	7 000	—
25–29	10 000	8 000	7 000	—
30–34	10 000	6 000	7 000	—
35–39	10 000	6 000	7 000	6 000
40–44	5 000	6 000	7 000	6 000
45–49	5 000	6 000	7 000	6 000
50–54	3 000	5 000	7 000	5 000
55–59	2 000	4 000	6 000	4 000
60–64	2 000	4 000	5 000	4 000
65–69	1 000	3 000	4 000	—
70–74	1 000	2 000	3 000	—
75–79	500	1 000	2 000	—
80–84	300	500	1 000	—
85 and over	200	500	1 000	—
Total	100 000	100 000	100 000	31 000

<sup>a</sup> From Waterhouse *et al.* (1976)



Table 2.5 Standard million population<sup>a</sup> of the USA in 1950 and 1970

Age range (years)	Standard million population	
	1950	1970
0-4	107 258	84 416
5-9	85 591	98 204
10-14	73 785	102 304
15-19	70 450	93 845
20-24	76 191	80 561
25-29	81 237	66 320
30-34	76 425	56 249
35-39	74 629	54 656
40-44	67 712	58 958
45-49	60 190	59 622
50-54	54 893	54 643
55-59	48 011	49 077
60-64	40 210	42 403
65-69	33 199	34 406
70-74	22 641	26 789
75-79	14 725	18 871
80-84	7 025	11 241
85+	3 828	7 435
Total	1 000 000	1 000 000

<sup>a</sup> From US Bureau of Censuses (1972)

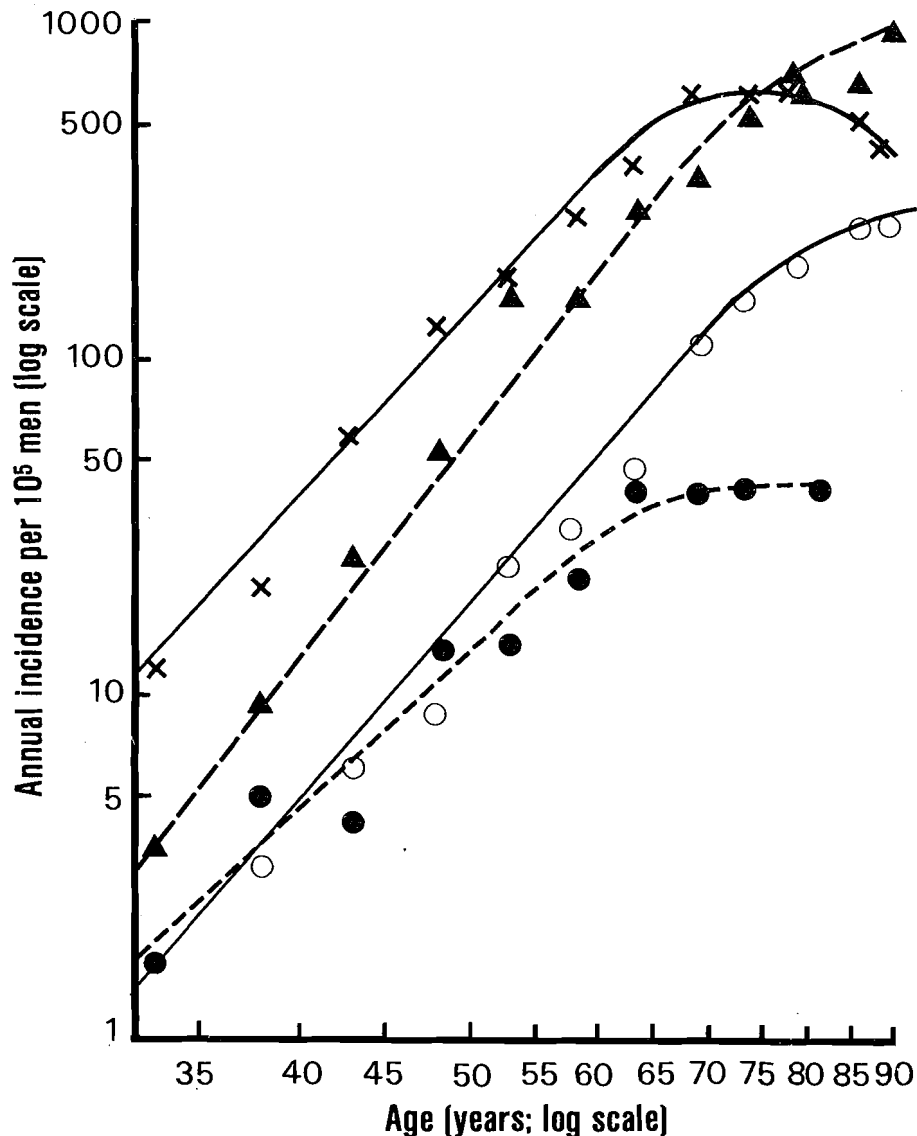
weights are used, one speaks of a rate that is 'standardized to the world population'. Alternatively, if the weights correspond to an actual age distribution, one would speak of a rate 'standardized to the population of the USA in 1950', for example, or of one standardized to 'the population of England and Wales in 1970'.

### Example 2.2

The crude death rates shown in the penultimate row of Table 2.3 steadily increase as a consequence of the general ageing of the cohort over time. In order to summarize the age-specific rates for different calendar periods, we calculated directly standardized rates for each one, using the 1950 US population (Table 2.5) as the standard. However, only rates for ages 40-79 were included in the calculation since the other age groups lacked data for one or more calendar periods. This necessity of discarding relevant data is one of the disadvantages of direct standardization. The standardized rates first rise and decline, as is true for most of their age-specific components. The initial rise is probably due to the 'healthy worker' selection bias (see §1.5c) which would apply to a large number of workers in the first calendar period, since everyone followed from the beginning of 1938 was still employed at that time. The eventual fall in the age-specific or age-standardized rates conforms to the pattern observed in the general population.

Provided that the same standard age distribution is used in their construction, comparison of directly standardized rates between different groups is thought to eliminate the differences that are observed in the crude rates solely by virtue of one group having a different age structure from another. However, as the graphs of cross-sectional age-incidence curves in Figure 2.2 make clear, such comparisons may

Fig. 2.2 Relationship between incidence of cancer of the stomach and age in four areas:  $\blacktriangle$ , Iceland;  $\times$ , Miyagi, Japan;  $\circ$ , Connecticut, USA;  $\bullet$ , Johannesburg & Kampala (African). From Doll and Cook (1967)



obscure important differences in the age-specific patterns. The apparent decline in stomach cancer incidence in older Japanese, in contrast to the rising age-incidence curve in Iceland, even among the elderly, means that the relative positions of the two countries as expressed in the age-standardized rate will depend to a large extent on the choice of the standard. If the standard population is heavily weighted towards the elderly, Iceland will have a relatively higher age standardized rate, while the reverse will be true if the standard population is younger. Doll and Cook (1967), from whose work the figure is taken, give several more examples of how the choice of the standard population affects the rank ordering of countries in terms of age-standardized incidence rates of specific cancers.

When incidence rates for cancers of epithelial tissues are plotted against age on a

log-log scale, they are often remarkably close to straight lines, with slopes of 4 or 5. Doll (1971) and others have interpreted this basic feature of incidence data as support for the concept that such cancers are produced by a series of cellular events. If there is curvature in the log-log plot, as in Figure 2.2, it is generally in the downward direction (Cook *et al.*, 1969). Sometimes this is due to a 'birth cohort effect', i.e., a general increase in rates for successive generations due presumably to the introduction of new agents into the personal or ambient environment. In this case, the curves flatten out or otherwise assume similar shapes when arranged to present age-specific data for successive generations of individuals born in the same time interval (birth cohorts) rather than for separate calendar periods. In other situations, possibly including the data shown in Figure 2.2, the curvature may represent the failure to diagnose completely incident cases among the elderly. Largely for this reason, Doll and Cook (1967) suggested that the calculation of summary rates for epithelial tumours be restricted to people aged 35–64, and introduced for this purpose the truncated population shown in the fourth column of Table 2.4. They argued that the directly standardized rate based on this population was a good measure of the average level of incidence or mortality, that the ratio of rates for the 60–64 *versus* the 35–39-year age groups measured the steepness of the increase with age, and that the two measures taken together provided a basic summary of the age-specific data.

#### (b) *The cumulative rate*

Cumulative rates are defined by equation (2.1) if one takes for  $w_j$  the length of the  $j$ th age interval rather than the standard age proportion (see §2.3 of Volume 1). Essentially the same measure was introduced by Yule (1934), except that he calculated an average of the age-specific rates rather than their sum, and termed the result the 'equivalent average death rate'. Since the nonzero weights of the truncated population are almost constant across the five-year age groups, a rate that is standardized in this fashion will be very nearly proportional to the cumulative rate between 35 and 64 years.

The cumulative rate has several advantages as a method of reporting cancer incidence and mortality data (Day, 1976). First it dispenses with the rather arbitrary selection of the standard population, yet has the desired feature of summarizing the age-specific data. Second, cumulative rates for different age ranges are additive. Thus, for example, the cumulative rate between 0 and 64 years is the sum of the cumulative rates for 0–34 and 35–64 years. Finally, the cumulative rate is easily converted into the cumulative risk by means of the formula  $P = 1 - \exp(-\Lambda)$ . This is the actuarial probability of disease development or death from the cause of interest, in the absence of other causes of death, for someone who is at risk throughout the designated age range. Since  $1 - \exp(-\Lambda)$  is approximately equal to  $\Lambda$  for small  $\Lambda$ , moreover, the cumulative rate can be roughly interpreted as the cumulative risk (actuarial probability), provided that it is small, say less than 10%. See Table 2.6.

#### **Example 2.3**

Table 2.7 compares cumulative incidence rates between 0 and 74 years to directly standardized rates based on two different standard populations (Day, 1976). It illustrates clearly that, while cancer is a 'rare' disease

Table 2.6 Conversion of cumulative rates ( $100\Lambda$ ) into the corresponding cumulative risks ( $100(1 - e^{-\Lambda})$ )

100 $\Lambda$	0.1	0.5	1.0	5.0	7.0	10.00	15.00	20.00	30.00
$100(1 - e^{-\Lambda})$	0.1	0.499	0.995	4.88	6.76	9.52	13.93	18.13	25.92

Table 2.7 Cumulative incidence rates, 0–74 years (percent), compared with rates standardized to the world population<sup>a</sup>, and to the truncated rate<sup>a</sup>, per 100 000 *per annum*

Population	Rate	Cancer site							
		Stomach	Lung	Breast	Cervix uteri	Leukaemia	Prostate	All sites <sup>b</sup>	
		Male	Male	Female	Female	Male	Male	Male	Female
Cali, Colombia	Cumulative (%)	7.34	2.14	3.08	8.35	0.38	2.71	29.11	29.78
	World population	57.5	17.5	27.3	75.6	5.2	23.2	(25.25)	(25.75)
	Truncated	87.7	29.0	62.6	183.6	5.6	2.15		
Alameda County, black	Cumulative (%)	2.69	5.85	4.15	3.18	1.05	7.54	30.80	20.81
	World population	24.4	43.8	38.6	30.5	8.3	65.3	(26.51)	(18.78)
	Truncated	33.0	88.9	75.0	75.4	8.9	55.6		
Birmingham, UK	Cumulative (%)	3.13	9.73	5.58	1.39	0.52	1.85	30.11	21.96
	World population	25.2	73.3	51.1	13.6	5.3	18.4	(26.00)	(19.50)
	Truncated	35.9	133.5	114.1	34.2	6.2	10.9		
Japan, Miyagi Prefecture	Cumulative (%)	11.97	2.16	1.06	2.28	0.36	0.35	24.22	16.30
	World population	95.3	15.6	11.0	20.6	4.4	3.2	(21.51)	(15.04)
	Truncated	164.1	22.6	27.5	52.8	4.7	2.0		

<sup>a</sup> From Doll *et al.* (1970)<sup>b</sup> The figures given in parentheses are the exact cumulative probabilities  $\hat{P} = 1 - \exp(-\hat{\Lambda})$  to compare with the cumulative incidence  $\hat{\Lambda}$ . See text and Table 2.6

when considered in terms of annual incidence, the total actuarial risk over a normal lifetime may be substantial. Japanese males, for example, have a cumulative actuarial risk for stomach cancer of 12%. Since the cumulative lifetime risk of many of the common cancers seen in laboratory animals is of the same order of magnitude (e.g., 5–40%), it is clear that expressing cancer incidence in such terms offers the possibility of more immediate extrapolation between epidemiology and laboratory investigations than does use of annual incidence rates (Peto, R. 1977).

### (c) Standard error of the cumulative or directly standardized rate

When death rates are computed from national vital statistics, or incidence rates are determined from cancer registries that cover large populations, questions of statistical or sampling stability are generally of minor importance. Errors inherent in the process of data collection, in the coding of cause of death or cancer type, or in the estimation of the population denominators are usually of much greater magnitude and concern. Rates calculated for study cohorts of limited size, however, may be based on a relatively small number of cases. Then, a simple formula for the standard error is useful as a measure of the statistical precision with which the rate is determined.

The formula given here stems from the elementary statistical model for the sampling distribution of a rate that is developed in some detail in §4.2. For the moment it suffices to know that the sampling variability of the rate numerator is approximately  $\text{Var}(d_j) = n_j \lambda_j$ , which may be estimated by  $d_j$  itself, and that the person-years denominators  $n_j$  may be regarded as fixed constants. Thus, the standard error of the age-specific rate  $\hat{\lambda}_j$ , i.e., the square root of the estimated variance, is  $\sqrt{d_j/n_j}$ .

In order to determine the variance or standard error of the summary rate  $\hat{\Lambda} = \sum w_j \hat{\lambda}_j$ , we need to know the covariances between the observed numbers of deaths in the different age intervals. The covariances are zero if the observations are statistically independent. This is true when the rates are estimated from cross-sectional data, for then different individuals are at risk in different age intervals. In cohort studies, however, the same person may contribute observation time to several contiguous age groups. Then, the  $d_j$  are not statistically independent, since the death of an individual in one interval precludes his dying in the next. Nevertheless, the discussion in §4.2 suggests that even in this case the  $d_j$  may be regarded as being independent for purposes of making large-sample statistical inferences. Chiang (1961) argues that the  $d_j$  are uncorrelated (see also Keyfitz, 1966). It follows that the standard error is

$$\text{SE}(\hat{\Lambda}) = \left( \sum_{j=1}^J w_j^2 \text{Var}(\hat{\lambda}_j) \right)^{1/2} = \left( \sum_{j=1}^J w_j^2 d_j / n_j \right)^{1/2}. \quad (2.2)$$

Inspection of equation (2.2) emphasizes another potential weakness of direct standardization, namely that the a-priori weights  $w_j$  take no account of the precision with which the component rates are estimated. The data for a single age interval may make a major contribution to the sampling error if the corresponding rate is based on a small denominator yet is given a large weight.

#### Example 2.4

We illustrate the calculation of cumulative rates and their standard errors by applying equations (2.1) and (2.2) to data from the Montana cohort. Table 2.8 shows the number of respiratory cancer deaths that occurred among the smelter workers at ages 40–79 in four calendar periods. The population denominators differ slightly from the corresponding entries in Table 2.2, since they were calculated according to another approximate method that is described in §3.1. Note that the 40–79-year age range accounts for  $276/288 = 96.5\%$  of the total deaths from this cancer. Since the age intervals are of equal length of ten years each, the formula for the standard error may be simplified to

$$\text{SE}(\hat{\Lambda}) = 10 \sqrt{\sum_{j=1}^J d_j / n_j}.$$

Cumulative respiratory cancer mortality rates between ages 40–79 are 8.40, 14.07, 13.81 and 14.41% for the four calendar periods. These may be compared with cumulative rates of 2.19, 4.21, 6.58 and 8.92% for the US white male population for the periods 1940–1949, 1950–1959, 1960–1969 and 1970–1975 (Appendix III). Thus the Montana cohort has substantially higher rates in the early decades, but the effect is attenuated somewhat by the passage of time. Part of the explanation for the decline in both relative and excess risk is that the later calendar years contain more person-years of observation from workers first employed after 1925, when the smelting process was changed and airborne exposures were presumably reduced (Lee-Feldstein, 1983).

US mortality rates for respiratory cancer are higher than those of the three western states near the smelter, namely Montana, Idaho and Wyoming. Use of standard rates from these states alone therefore increases the discrepancy between the respiratory cancer rates for the cohort and those for the surrounding general

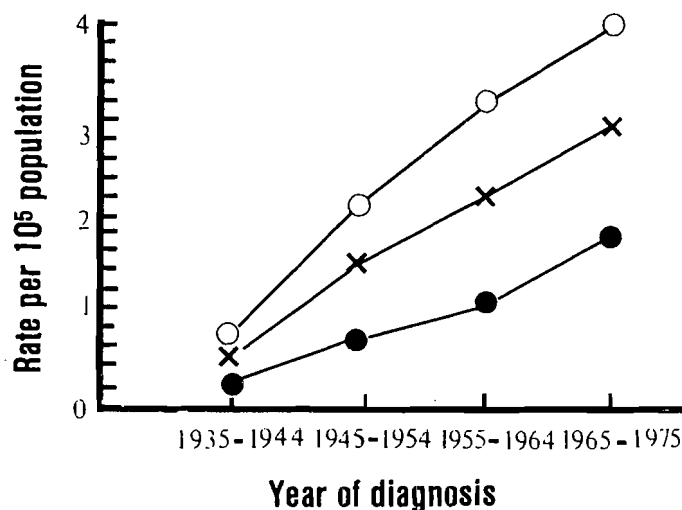
Table 2.8 Respiratory cancer deaths ( $d$ ), person-years at risk ( $n$ , in thousands), and death rates ( $\hat{\lambda}$ , per 1000 person-years) in the Montana cohort. Calculation of the cumulative rate and its standard error

Age range (years)	Calendar period				Totals	
	1938–1949	1950–1959	1960–1969	1970–1977		
40–49	<i>d</i>	5	5	7	4	21
	<i>n</i>	9.217	14.949	16.123	9.073	49.363
	$\hat{\lambda}$	0.542	0.334	0.434	0.441	0.425
50–59	<i>d</i>	11	24	28	17	80
	<i>n</i>	6.421	10.223	13.663	11.504	41.811
	$\hat{\lambda}$	1.713	2.348	2.049	1.478	1.913
60–69	<i>d</i>	14	24	44	35	117
	<i>n</i>	4.006	4.896	7.555	7.937	24.394
	$\hat{\lambda}$	3.495	4.902	5.824	4.410	4.796
70–79	<i>d</i>	4	12	15	27	58
	<i>n</i>	1.507	1.851	2.724	3.341	9.423
	$\hat{\lambda}$	2.654	6.483	5.506	8.081	6.155
Totals	<i>d</i>	34	65	94	83	276
	<i>n</i>	21.151	31.920	40.066	31.855	124.991
	$\hat{\lambda}$	1.608	2.036	2.346	2.606	2.208
Cumulative rate (%)		8.405	14.067	13.814	14.410	13.290
Standard error (%)		1.720	2.181	1.723	1.775	0.951

population. Regional rather than national death rates generally make a more appropriate standard, but they are often not available for the entire time period of interest or are based on such small populations as to be unstable.

Our next example confirms the basic point that age standardization techniques as discussed in this section can obscure important features of the data and should be used cautiously in analytical work.

Fig. 2.3 Thyroid cancer incidence rates, 1935–1975, for Connecticut, USA, age-adjusted to the 1950 US population: ○, females; ●, males; ×, both sexes. From Mendelsohn–Pottern *et al.* (1980)



**Example 2.5**

Figure 2.3 shows age-adjusted incidence rates for thyroid cancer from the Connecticut Tumor Registry by sex and ten-year calendar period from 1935–1975 (Mendelsohn-Pottern *et al.*, 1980). They were calculated by direct standardization relative to the 1950 US population (Table 2.5). While they show a smoothly rising incidence over the 40-year period, they miss an important feature of the data for females. When the age-specific rates are plotted for each of the four periods (Figure 2.4), a clear bimodal age-incidence curve emerges for females after 1954, with a first peak between 25 and 44 years of age and a steady increase in rates from age 65 on. The first peak was less pronounced in males. It was discovered to be due to increases in rates for papillary and follicular carcinomas and was interpreted as probably due to increased childhood exposure to therapeutic radiation.

**2.3 Comparative measures of incidence and mortality**

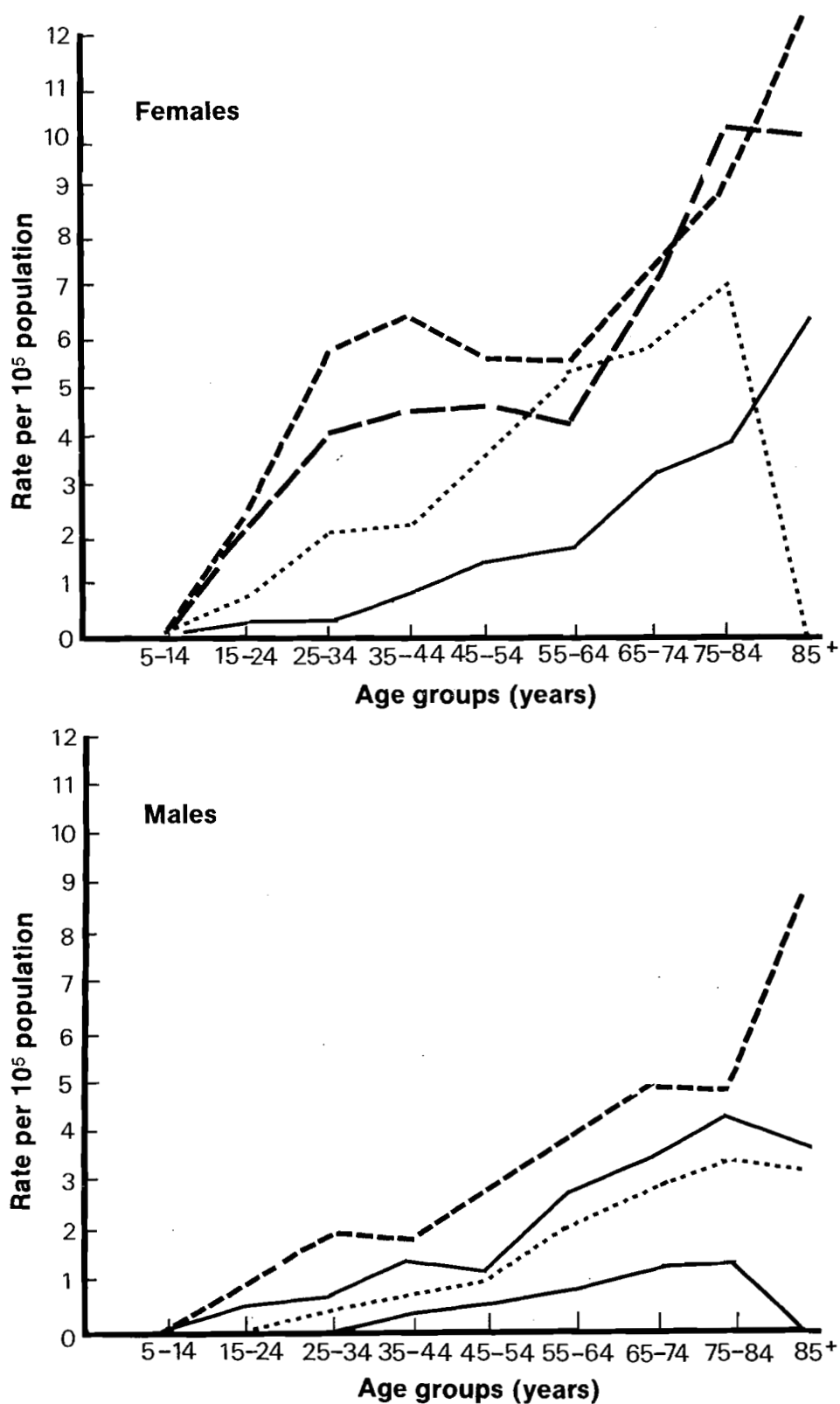
A major goal of standardization, besides combining a set of age-/stratum-specific rates into a synoptic figure, is to provide a quantitative measure of the difference in rates between the study cohort and a standard population or other comparison group that is free from the effects of age or other confounding variables. Rather than taking the ratio or difference of the crude mortality rates for cohort *versus* standard as a measure of effect, one first divides the comparison groups into a number of strata that are reasonably homogeneous with respect to the confounding variables. The stratum-specific rates for both groups are calculated and their differences or ratios are summarized in a single comparative figure. Since ratios of age-specific cancer incidence rates are typically more nearly constant than are the rate differences (see §2.5 of Volume 1), a summary ratio is generally the more appropriate measure. However, caution must be exercised if, as in the last example, there are substantial variations between the comparison groups in the age-specific ratios. In such circumstances the investigator is better advised to choose some other measure of effect (such as a rate difference) that does remain constant, or else to model the variations in rate ratios or rate differences as a function of age and other stratification variables, rather than attempting to summarize them in a single number.

The choice of variables to be used as a basis for stratification or other statistical adjustment procedure raises several complicated issues (see §3.4 of Volume 1). One generally wants to adjust for variables that are causally related to disease, and the differential distribution of which among the comparison groups could therefore result in apparent differences in incidence or mortality that are secondary to the causal effects. This implies that some prior understanding or hypotheses about the causal nature of the disease process necessarily enters into the selection of stratification variables. Questions of the statistical significance, in the data under study, of their association with either the disease or the exposures are secondary if not irrelevant. Age is the paradigm case of a confounding variable since it is usually regarded as an independent cause, or at least as a surrogate for the accumulation of independent causes, of many cancers and other diseases.

**(a) The comparative mortality figure (CMF)**

A simple summary of the incidence or mortality rate ratios between the cohort and standard population that accounts for the possible confounding effects of age or other

Fig. 2.4 Age-specific thyroid cancer incidence rates, 1935–1975, for Connecticut, USA: ····, 1965–1975; —, 1955–1964; ----, 1945–1954; ———, 1935–1944. From Mendelsohn-Pottern *et al.* (1980)





variables is obtained by dividing the directly standardized rate for the cohort by the standard population rate. Thus if  $\lambda_j^*$  denotes the standard death rate in stratum  $j$ , and  $w_j$  is the standard weight, the comparative mortality figure (CMF) is defined by

$$\text{CMF} = \frac{\sum_{j=1}^J w_j d_j / n_j}{\sum_{j=1}^J w_j \lambda_j^*}. \quad (2.3)$$

The ratio of CMFs calculated for two different cohorts using the same standard rates and weights is simply the ratio of the two directly standardized rates.

When we introduced the concept of direct standardization, the standard weights  $w_j$  were chosen to be equal to the person-years denominators  $n_j^*$  of the standard rates. With these weights one may write

$$\text{CMF} = \frac{\sum w_j d_j / n_j}{\sum w_j d_j^* / n_j^*} = \frac{\sum n_j^* d_j / n_j}{D^*}, \quad (2.4)$$

where  $d_j^*$  and  $n_j^*$  denote standard deaths and person-years in stratum  $j$  and  $D^* = \sum d_j^*$  represents the total standard deaths. The second expression is easier for computation. An interpretation of the CMF in this case is as the ratio of the number of deaths that would be expected in the cohort if it had the same age structure as the standard population, using the stratum-specific cohort rates to calculate the expectation, divided by the number of deaths in the standard population. This version of the CMF may also be recognized as a weighted average of age-specific cohort to standard rate ratios  $r_j = \hat{\lambda}_j / \lambda_j^*$ ,

$$\text{CMF} = \frac{\sum_{j=1}^J u_j r_j}{\sum_{j=1}^J u_j}, \quad (2.5)$$

where now the weights  $u_j = n_j^* \lambda_j^* = d_j^*$  are equal to the number of deaths in each age group in the standard population.

A major disadvantage of the CMF is its instability when the component rates are based on small numbers of deaths. This is easily illustrated by means of a hypothetical example.

### Example 2.6

Table 2.9, adapted from Mosteller and Tukey (1977, p. 237), presents fictitious data involving three age strata. The CMF is determined from equation (2.4) as

$$\text{CMF} = \frac{150\,000(10/10\,000) + 70\,000(9/3000) + 210(1/1)}{460} = 1.24.$$

However, if the single cohort member in the 85+ age stratum were to survive instead of die, the same calculation gives

$$\text{CMF} = \frac{150\,000(10/10\,000) + 70\,000(9/3000) + 210(0/1)}{460} = 0.78.$$

Thus, a change in only one of the 46 deaths has made a large difference in the comparative analysis.

Table 2.9 Fictitious data used to illustrate the instability of the CMF<sup>a</sup>

Age stratum (years)	Cohort		Standard population	
	Deaths (d)	Person-years (n)	Deaths (d)	Person-years (n)
45–64	10	10 000	140	150 000
65–84	9	3 000	290	70 000
85+	1	1	30	210
Totals	20	13 001	460	220 210

<sup>a</sup> Adapted from Mosteller and Tukey (1977)*(b) Standard error of the CMF*

Instead of examining its sensitivity to individual deaths, a more systematic way of measuring the statistical precision of the CMF is to calculate its standard error (SE). We assume that the standard population is very large relative to the cohort, so that sampling errors in the standard rates may be ignored. Then, the standard error of the CMF is obtained directly from the standard error of its numerator. From equations (2.2) and (2.3) we have

$$SE(\text{CMF}) = \frac{(\sum_{j=1}^J w_j^2 d_j / n_j^2)^{1/2}}{\sum_{j=1}^J w_j \lambda_j^*}. \quad (2.6)$$

Corrections to this equation are needed if the standard population is constructed as a pool of several cohorts that includes the one for which the CMF is being determined (Yule, 1934).

If the standard error is not regarded simply as a measure of statistical precision, but is to be used to construct test statistics or confidence intervals, it is preferable to make a transformation to the log scale. This helps to correct the skewness in the statistical distribution of the CMF itself and thus improves the normal approximation to the distribution of test statistics based on it. The standard error of the transformed CMF is

$$SE(\log \text{CMF}) = \frac{SE(\text{CMF})}{\text{CMF}} = \frac{(\sum_{j=1}^J w_j^2 d_j / n_j^2)^{1/2}}{\sum_{j=1}^J w_j d_j / n_j}. \quad (2.7)$$

A test of the null hypothesis  $\text{CMF} = 1$  could be effected by referring  $\log \text{CMF} / SE(\log \text{CMF})$  to tables of the normal distribution, but, in practice, such tests are better carried out directly in terms of the age-specific rates, as described in the next chapter.

Similar considerations apply to a comparison of the CMFs for two different cohorts, or equivalently to a comparison of the corresponding directly standardized rates. The standard error of the log ratio  $\log(\text{CMF}_2 / \text{CMF}_1) = \log \text{CMF}_2 - \log \text{CMF}_1$  is given by

$$SE \log \frac{\text{CMF}_2}{\text{CMF}_1} = \left\{ \frac{\sum_{j=1}^J w_j^2 d_{1j} / n_{1j}^2}{(\sum_{j=1}^J w_j d_{1j} / n_{1j})^2} + \frac{\sum_{j=1}^J w_j^2 d_{2j} / n_{2j}^2}{(\sum_{j=1}^J w_j d_{2j} / n_{2j})^2} \right\}^{1/2}.$$

(c) *The standardized mortality ratio (SMR)*

The standardized mortality ratio (SMR) has been in service at least since 1786 (Keiding, 1987). It was used by W.H. Farr in the 1855 annual report of the Registrar General of Great Britain to compare mortality in different occupational groups (Benjamin, 1968). It is also defined as a weighted average of the age-specific rate ratios (equation 2.5), where the weights  $w_j = n_j \lambda_j^*$  are the expected number of deaths for the cohort in the  $j$ th age group, rather than the number of standard deaths as used for the CMF. Thus, the SMR compares the observed number of deaths in the cohort with an expected number obtained by applying the standard rates to the cohort age structure. In symbols,

$$\text{SMR} = \frac{\sum_{j=1}^J d_j}{\sum_{j=1}^J n_j \lambda_j^*} = \frac{D}{E^*}, \quad (2.8)$$

where  $D = \sum d_j$  denotes the total observed number of deaths and  $E^*$  is the expected number.

In typical applications, the SMR is used to compare mortality from each of several causes of death in the cohort as a whole to that in the general population. Table 2.16, for example, shows the SMRs for four causes of death for the Montana smelter workers. Diseases identified as occurring in excess may then be studied in greater detail in relation to specific exposures. Of course, there is no guarantee that this process will identify those diseases or causes of death that are most closely associated with the exposures. Cause-specific rates for unexposed cohort members may be less than those in the general population, whereas rates for exposed members are higher, and the two effects may cancel each other out when averaged over the entire cohort. Nevertheless, use of the SMR as a device for screening a number of different causes of death seems firmly established. Other techniques to detect cancer sites or causes of death that are related to exposure, but without reference to an external standard population, are considered in Chapter 3.

One advantage of the SMR over the CMF is that age-specific numbers of deaths  $d_j$  are not required for its calculation. It suffices to know only the total  $D$ . This sometimes permits application of the SMR to published data for which the CMF could not be used. Details on numbers of deaths by cause, subgroup and age are often omitted from official publications for reasons of economy, whereas the subtotals by cause and subgroup and the person-years by subgroup and age are given. However, caution is required in such circumstances, because if the detailed data are missing there is no way of evaluating the hypothesis of constant rate ratios that is needed to justify fully the use of these summary measures (see below and also §4.6).

The SMR is also the preferred measure when analysing cross-sectional data according to birth cohort rather than calendar period. The reason is that the age intervals for which data are available differ for different generations, a feature that precludes calculation of comparable CMFs. Beral (1974) and Beral *et al.* (1978) have provided us with two particularly innovative examples that illustrate this type of application.

**Example 2.7**

Figure 2.5 shows SMRs for ovarian cancers calculated for successive generations of women from age-specific mortality rates. Data were available for women aged 30–74 years for the period 1931 to 1973 in the USA and for the same ages between 1931 and 1975 in the UK. Age-specific rates for the pooled calendar periods were used as a standard, and the SMRs consequently tend to cluster about 100. When the SMRs are plotted against average completed family size for the same generations, there is a near perfect negative correlation that suggests a possible protective effect of pregnancy or childbearing (Fig. 2.6).

In a similar analysis, Figure 2.7 shows plots of SMRs for cervical cancer in England and Wales by generation together with the rates of gonorrheal disease that prevailed at age 20 in those same generations. The similarity in the shapes of the two curves is striking.

Another advantage of the SMR, when viewed as a weighted average of the ratios of age-specific rates for cohort and standard population, is that the weights  $w_j = n_j \lambda_j^*$  minimize the variance of the weighted average. Assuming that the true rate ratios are constant, the SMR is thus the minimum variance estimate of the common rate ratio. In practical terms, this means that it tends to be less sensitive to numerical instabilities in one or two of the age-specific rates, a property that is easy to demonstrate by returning to an earlier example.

**Example 2.6 (contd)**

The expected numbers of deaths for the cohort in Table 2.9 is

$$E^* = 10\,000(140/150\,000) + 3000(290/70\,000) + 1(30/210) = 21.90,$$

and thus the SMR is  $20/21.9 = 0.91$ , indicating a slightly lower death rate among the cohort members as opposed to the general population. If the single exposed person in the 85+ age group had died instead of lived, we would have  $\text{SMR} = 19/21.90 = 0.87$ , a minor change compared to that observed earlier with the CMF.

Another means of demonstrating the greater numerical stability of the SMR is to

Fig. 2.5 Age-standardized mortality ratios for ovarian cancer in England and Wales (●) and the USA (○) for generations of women born at five-year intervals between 1861 and 1931. From Beral *et al.* (1978)

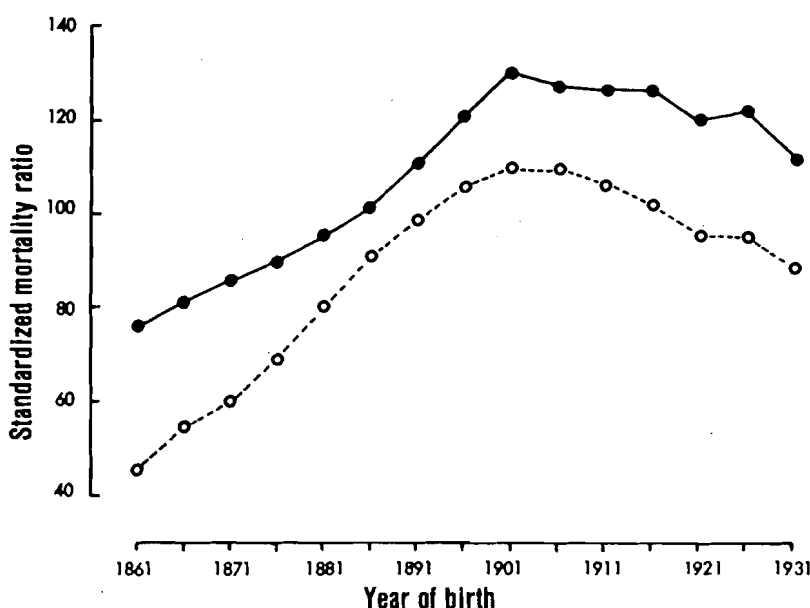
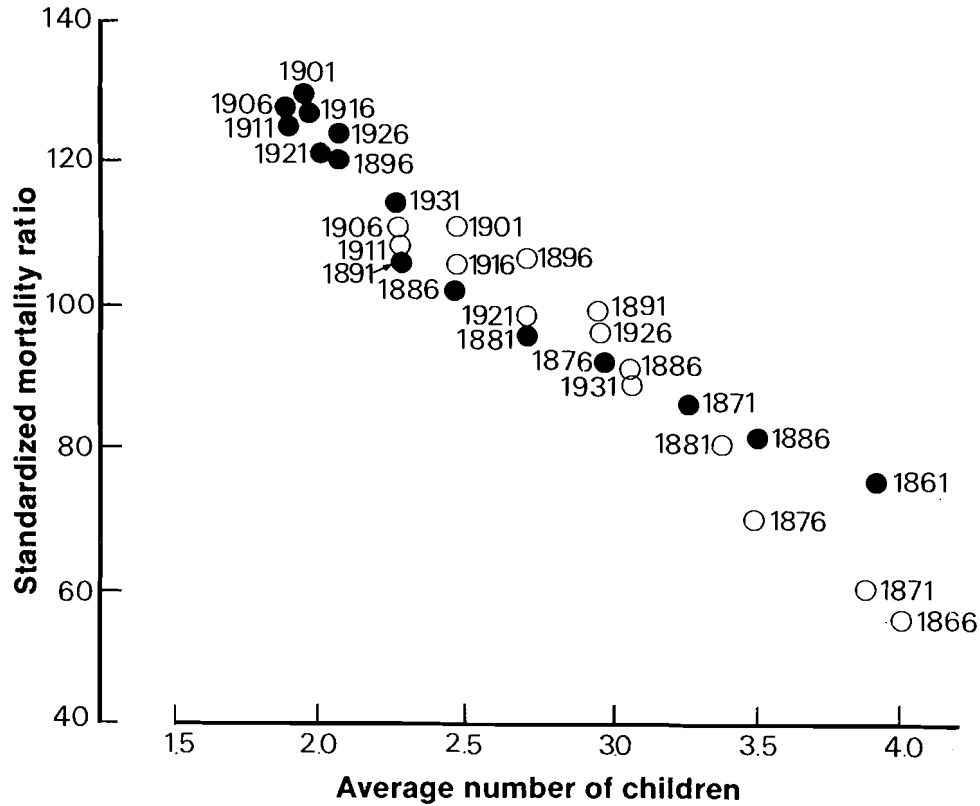


Fig. 2.6 Age-standardized mortality ratios for ovarian cancer plotted against the average completed family size (number of children) for different generations of women in England and Wales (●) and the USA (○). The mid-year of birth of each generation is shown in parentheses. From Beral *et al.* (1978)



examine its standard error. Under the previous assumption that the numbers of deaths  $d_j$  in the different (age) strata are uncorrelated, whereas sampling errors associated with the standard population are negligible, we calculate

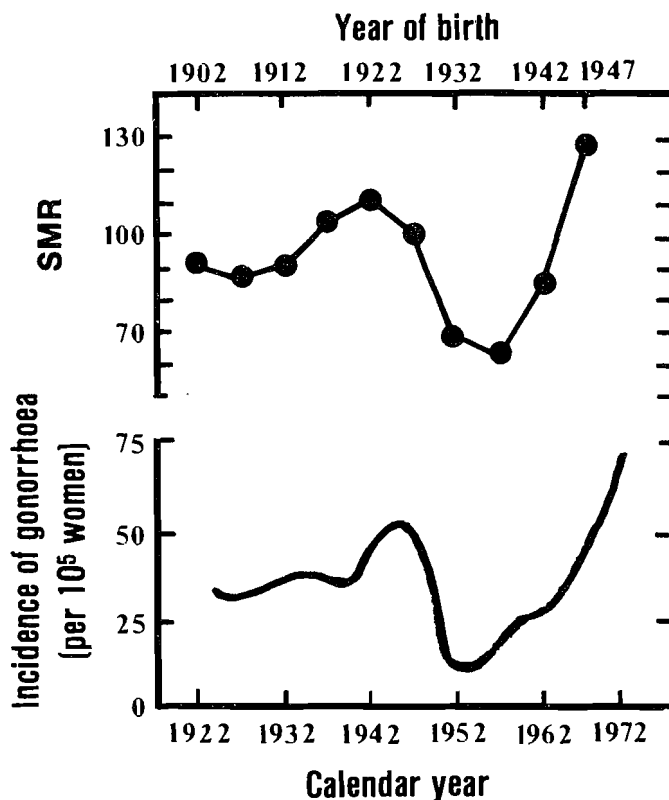
$$SE(SMR) = \left\{ \sum_{j=1}^J d_j \right\}^{1/2} / E^* = D^{1/2} / E^*,$$

or, more appropriately

$$SE(\log SMR) = SE(SMR) / SMR = 1 / D^{1/2}. \quad (2.9)$$

Since the standard error of the SMR depends only on fluctuations in the total number rather than in the age-specific numbers of deaths, it is generally smaller than that of the CMF. As already noted, under the hypothesis of constant age-specific rate ratios, the SMR weights the ratios optimally, in inverse proportion to their statistical precision, whereas with the CMF the weights associated with unstable ratios may be much larger. The SMR is thus more appropriate when the sample is small and questions of statistical significance are at issue; we examine methods of making statistical inferences about this measure in more detail than we did for the CMF.

Fig. 2.7 Standardized mortality ratios (SMR) from cervical cancer by birth cohort among women born between 1902 and 1947 in England and Wales and incidence of gonorrhoea among women in England and Wales, 1925–1972. From Beral (1974)



(d) *Testing the significance of the observed SMR*

The first question of interest relating to the SMR is simply whether the observed cause-specific mortality in the study cohort can be explained adequately by the standard rates and the play of chance. Conventional approaches (Monson, 1980) use the simple continuity corrected chi-square statistic

$$\chi^2 = \frac{(|D - E^*| - 1/2)^2}{E^*} \quad (2.10)$$

in order to test whether the observed number of deaths is significantly different from the number expected. This statistic is derived from the usual assumption that, under the null hypothesis, the observed number of deaths  $D$  is approximately Poisson distributed with mean and variance both equal to  $E^*$  (Armitage, 1971, section 4.3). It is referred to tables of chi-square with one degree of freedom, or else its (signed) square root  $\chi$  is treated as an equivalent normal deviate. The  $1/2$  correction in the numerator is intended to improve the correspondence between the percentiles of the discrete Poisson distribution and the continuous normal one (see §4.3 of Volume 1).

When the number of deaths is small, the Poisson distribution is rather skewed, and

the normal approximation implicit in the use of (2.10) will be inadequate. An 'exact'  $p$  value may be calculated using tail probabilities of the Poisson or (equivalently) chi-square distributions (Pearson & Hartley, 1966). However, these are tabled for only a limited range of values. Byar (see Rothman & Boice, 1979) suggested an extremely accurate approximation to the exact Poisson test, which is obtained by calculating the deviate

$$\chi = (9\tilde{D})^{1/2} \left\{ 1 - \frac{1}{9\tilde{D}} - \left( \frac{E^*}{\tilde{D}} \right)^{1/3} \right\}, \quad (2.11)$$

where  $\tilde{D} = D$  if  $D$  exceeds  $E^*$ , and  $\tilde{D} = D + 1$  otherwise, and referring it to tables of the unit normal distribution (Rothman & Boice, 1979). Alternatively, and somewhat easier to remember, we may use the fact that the square-root transform is 'variance stabilizing' (Armitage, 1971), so that  $D^{1/2}$  is approximately normal with mean  $(E^*)^{1/2}$  and variance  $1/4$  under the null hypothesis. This means treating

$$\chi = 2\{D^{1/2} - (E^*)^{1/2}\} \quad (2.12)$$

as a standard normal deviate.

#### (e) Confidence intervals for the SMR

A second statistical question of common interest is to determine a range of possible values for the true SMR that are reasonably consistent with the observed data. If the test of the null hypothesis gives the verdict 'not significant', it may be important to demonstrate that the study had sufficient precision to render large differences between cohort and standard rates implausible. Or, if the result is positive, one may wish to examine its consistency with that of other studies. Putting a confidence interval around the observed SMR can achieve these goals.

Exact confidence limits for the SMR are obtained by first finding lower ( $L$ ) and upper ( $U$ ) limits  $\mu_L$  and  $\mu_U$  for the mean  $\mu = E(D)$  of the Poisson distributed observation  $D$ , and then calculating  $SMR_L = \mu_L/E^*$  and  $SMR_U = \mu_U/E^*$ . Exploiting the general relationship between confidence limits and test statistics (§§4.2 and 4.3 of Volume 1), the limits on  $\mu$  may be found by solution of equations involving Poisson probabilities. Table 2.10 presents exact 95% limits for a Poisson mean for selected values of  $D$  ranging from 1 to 1000 (Haenszel *et al.*, 1962).

For other confidence levels and values of  $D$  not shown in Table 2.10, Byar's approximation is sufficiently accurate that one may avoid the iterative calculations needed for the exact results. Thus, for a  $100(1 - \alpha)\%$  confidence interval, we have the approximate limits (Rothman & Boice, 1979)

$$\mu_L = D \left( 1 - \frac{1}{9D} - \frac{Z_{\alpha/2}}{3D^{1/2}} \right)^3 \quad (2.13)$$

and

$$\mu_U = (D + 1) \left( 1 - \frac{1}{9(D + 1)} + \frac{Z_{\alpha/2}}{3(D + 1)^{1/2}} \right)^3,$$

where  $Z_{\alpha/2}$  denotes the  $100(1 - \alpha/2)$  percentile of the unit normal distribution.

Table 2.10 Tabulated values of 95% confidence limit factors for estimating a Poisson-distributed variable<sup>a</sup>

Observed number on which estimate is based (n)	Lower limit factor (L)	Upper limit factor (U)	Observed number on which estimate is based (n)	Lower limit factor (L)	Upper limit factor (U)	Observed number on which estimate is based (n)	Lower limit factor (L)	Upper limit factor (U)
1	0.0253	5.57	21	0.619	1.53	120	0.833	1.200
2	0.121	3.61	22	0.627	1.51	140	0.844	1.184
3	0.206	2.92	23	0.634	1.50	160	0.854	1.171
4	0.272	2.56	24	0.641	1.48	180	0.862	1.160
5	0.324	2.33	25	0.647	1.48	200	0.868	1.151
6	0.367	2.18	26	0.653	1.47	250	0.882	1.134
7	0.401	2.06	27	0.659	1.46	300	0.892	1.121
8	0.431	1.97	28	0.665	1.45	350	0.899	1.112
9	0.458	1.90	29	0.670	1.44	400	0.906	1.104
10	0.480	1.84	30	0.675	1.43	450	0.911	1.098
11	0.499	1.79	35	0.697	1.39	500	0.915	1.093
12	0.517	1.75	40	0.714	1.36	600	0.922	1.084
13	0.532	1.71	45	0.729	1.34	700	0.928	1.078
14	0.546	1.68	50	0.742	1.32	800	0.932	1.072
15	0.560	1.65	60	0.770	1.30	900	0.936	1.068
16	0.572	1.62	70	0.785	1.27	1 000	0.939	1.064
17	0.583	1.60	80	0.798	1.25			
18	0.593	1.58	90	0.809	1.24			
19	0.602	1.56	100	0.818	1.22			
20	0.611	1.54						

<sup>a</sup> From Haenszel *et al.* (1962)

Somewhat less accurate but more easily remembered approximate limits for the SMR may be derived from analogues to the other statistics (2.10) and (2.12) used to test the null hypothesis. Specifically, denoting by  $\theta$  the unknown value of the SMR, we solve the equation  $(D - \theta E)^2 / \theta E = Z_{\alpha/2}^2$  (ignoring the continuity correction) to find

$$\text{SMR}_L = \theta_L = \text{SMR} \left[ 1 + \frac{1}{2D} Z_{\alpha/2}^2 \{1 - (1 + 4D/Z_{\alpha/2}^2)^{1/2}\} \right] \quad (2.14)$$

and

$$\text{SMR}_U = \theta_U = \text{SMR} \left[ 1 + \frac{1}{2D} Z_{\alpha/2}^2 \{1 + (1 + 4D/Z_{\alpha/2}^2)^{1/2}\} \right]$$

as the limits derived from the standard chi-square test. We have not used a continuity correction for this calculation, since to do so gives less accurate limits empirically. Alternatively, limits based on the square-root transform are obtained by solving the equation  $2\{D^{1/2} - (\theta E)^{1/2}\} = Z_{\alpha/2}$ , which gives

$$\text{SMR}_L = \text{SMR} \left( 1 - \frac{Z_{\alpha/2}}{2D^{1/2}} \right)^2$$



and

$$SMR_U = SMR \left( \frac{D+1}{D} \right) \left( 1 + \frac{Z_\alpha}{2(D+1)^{1/2}} \right)^2. \quad (2.15)$$

The use of  $D+1$  in the second equation is made on strictly empirical grounds in order to improve the approximation for small  $D$  (compare equation (2.13)).

The exact Poisson limits and all three sets of approximate confidence limits (2.13)–(2.15) can be expressed in the form  $SMR_L = SMR \times M_L$  and  $SMR_U = SMR \times M_U$ , where  $M_L$  and  $M_U$  are multipliers determined by  $\alpha$  and  $D$ . Table 2.11 compares the multipliers obtained by each method for several values of  $D$ ,  $\alpha = 0.05$  and  $\alpha = 0.01$  (95% and 99% confidence). Byar's approximation is shown to be accurate even for quite small numbers of deaths. The square-root transform performs reasonably well as soon as  $D$  exceeds 10 or so. Especially as regards the lower bound, which is usually of prime interest, however, the approximation based on the simple chi-square statistic is not very satisfactory.

Table 2.11 Exact and approximate multipliers for computing confidence intervals for the SMR<sup>a</sup>

No. of deaths ( $D$ )	Exact limits		Byar's approximation (equation 2.13)		Square root (equation 2.15)		Chi-square (equation 2.14)	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
95% intervals								
1	0.025	5.572	0.013	5.565	0.000	5.733	0.176	5.667
2	0.121	3.612	0.112	3.611	0.094	3.678	0.274	3.647
3	0.206	2.922	0.201	2.922	0.188	2.961	0.340	2.941
4	0.272	2.560	0.269	2.561	0.260	2.586	0.389	2.572
5	0.325	2.334	0.322	2.334	0.315	2.353	0.427	2.342
10	0.480	1.839	0.479	1.839	0.476	1.846	0.543	1.841
15	0.560	1.649	0.559	1.650	0.558	1.653	0.606	1.650
20	0.611	1.544	0.611	1.545	0.610	1.547	0.647	1.545
25	0.647	1.476	0.647	1.476	0.646	1.478	0.677	1.476
50	0.742	1.318	0.742	1.318	0.742	1.319	0.759	1.318
99% intervals								
1	0.005	7.430	0.000	7.471	0.000	7.301	0.117	8.519
2	0.052	4.637	0.038	4.656	0.008	4.561	0.195	5.123
3	0.113	3.659	0.102	3.671	0.066	3.604	0.253	3.960
4	0.168	3.149	0.160	3.157	0.127	3.105	0.297	3.362
5	0.216	2.830	0.209	2.836	0.180	2.794	0.334	2.993
10	0.372	2.140	0.369	2.142	0.351	2.120	0.452	2.212
15	0.460	1.878	0.458	1.879	0.445	1.864	0.520	1.922
20	0.518	1.733	0.517	1.734	0.507	1.723	0.566	1.765
25	0.560	1.640	0.559	1.641	0.551	1.632	0.601	1.665
50	0.673	1.425	0.673	1.426	0.669	1.421	0.696	1.437

<sup>a</sup> Note: In order to obtain lower and upper limits for an SMR based on the indicated number of deaths, the computed SMR is multiplied by the values shown.

**Example 2.8**

Fifteen deaths from bladder cancer were recorded prior to 1964 among the Montana smelter workers, whereas only 8.33 were expected from US population rates:  $SMR = 15/8.33 = 1.80$ . The 95% confidence limits found from the exact multipliers corresponding to  $D = 15$  in Table 2.10 are  $SMR_L = 0.560 \times 1.80 = 1.01$  and  $SMR_U = 1.649 \times 1.80 = 2.97$ . Those based on Byar's formula are almost the same, and those for the square-root transform only very slightly wider (1.00, 2.98). However, the limits based on the chi-square test statistic (1.11, 2.97) have a lower limit which is seriously in error, as do those based on the standard error of  $\log(SMR)$ , namely  $1.80 \times \exp(\pm 1.96/\sqrt{15}) = (1.09, 2.99)$ .

Since the exact lower limit just excludes 1.0, we know that the exact two-sided significance level must be just under 0.05. Equation (2.11) gives  $\chi = 1.98$  ( $p = 0.048$ ), whereas with (2.12) we find  $\chi = 1.97$  ( $p = 0.049$ ). The conventional formula (2.10) yields  $\chi = 2.14$  ( $p = 0.033$ ) with continuity correction and  $\chi = 2.31$  ( $p = 0.021$ ) without. This reinforces our conclusion that the test statistics (2.11) and (2.12) should be used in preference to (2.10).

**(f) SMR versus CMF: a tradeoff between bias and variance**

Up until now we have emphasized the statistical advantages of the SMR over the CMF, but, unfortunately, this is not the entire story. The major weakness of the SMR is that ratios of SMRs for two comparison groups may not adequately represent the ensemble of ratios of their component age- or stratum-specific rates (Yule, 1934). In fact, as the schema shown in Table 2.12 makes clear, there is a precise analogy with the arithmetic of statistical confounding.

The ratios of SMRs for Cohort 1 *versus* Cohort 2 within each age group equal the odds ratio calculated from the corresponding  $2 \times 2$  table, and likewise the overall SMR ratio is the odds ratio from the totals table. According to the general principle of statistical confounding (§3.4 of Volume 1), it follows that, even if the two age-specific odds ratios are equal, they may differ from the pooled odds ratio if both (i) the SMRs for each cohort vary from one age group to another and (ii) the age distributions of the two cohorts, and hence the distributions of expected numbers of deaths, are disparate. Since the age-specific ratios of SMRs equal the ratios of the corresponding rates (assuming that the standard rates are used in calculation of the expected numbers), it follows that the ratio of two SMRs determined by pooling observed and expected deaths across age groups may sometimes lie completely outside the range of the age-specific rate ratios.

The fictitious data in Table 2.13 provide a clear numerical illustration of this phenomenon. The overall SMR for each cohort is a weighted average of the two

Table 2.12 Confounding remaining after indirect standardization

	Age group 1		Age group 2		Total	
	Observed	Expected	Observed	Expected	Observed	Expected
Cohort 1	$d_{11}$	$e_{11}^*$	$d_{12}$	$e_{12}^*$	$D_1$	$E_1^*$
Cohort 2	$d_{21}$	$e_{21}^*$	$d_{22}$	$e_{22}^*$	$D_2$	$E_2^*$
$SMR_1$	$\frac{d_{11}e_{21}^*}{d_{21}e_{11}^*}$		$\frac{d_{12}e_{22}^*}{d_{22}e_{12}^*}$		$\frac{D_1E_2^*}{D_2E_1^*}$	
$SMR_2$	$\frac{d_{21}e_{11}^*}{d_{11}e_{21}^*}$		$\frac{d_{22}e_{12}^*}{d_{12}e_{22}^*}$		$\frac{D_2E_1^*}{D_1E_2^*}$	

$d$  and  $D$  = number of deaths observed  
 $e^*$  and  $E^*$  = number of deaths expected

Table 2.13 Example of a misleading ratio of SMRs<sup>a</sup>

Cohort		Age range (years)		
		20–44	45–64	Total (20–64)
I	<i>d</i>	100	1600	1700
	<i>e</i> *	200	800	1000
	SMR <sub>1</sub> (%)	50	200	170
II	<i>d</i>	80	180	260
	<i>e</i> *	120	60	180
	SMR <sub>2</sub> (%)	67	300	144
	SMR <sub>1</sub> /SMR <sub>2</sub>	75	67	118

<sup>a</sup> Adapted from Kilpatrick (1963)*d* = number of deaths observed*e*\* = number of deaths expected

age-specific  $d/e^*$  ratios, the weights being proportional to the expected number of deaths. Since Cohort 1 has more older people, the high  $d/e^*$  ratio for the 45–64-year age interval is weighted more heavily, whereas in Cohort 2 much more emphasis is given to the lower  $d/e^*$  ratio in the 25–44-year age interval. The overall result is a change of sign in the apparent effect, from an excess of deaths in Cohort 2 on an age-specific basis to an apparent excess in Cohort 1 when the data are pooled.

The CMF does not suffer from this problem. Ratios of two CMFs, being ratios of directly standardized rates, can be expressed as a weighted average of the age-specific rate ratios. If these are all equal to some constant value  $\theta$ , therefore, the ratio of CMFs must also equal  $\theta$ . However, this equality does not hold for the SMRs, unless, in addition, the age-specific rates for each comparison group are also in constant proportion with those for the standard population. This bias in the SMR has led many authors to conclude that the CMF is the preferred measure. Miettinen (1972) says of the SMR that ‘estimates computed in this manner are internally standardized but not mutually comparable’. Kilpatrick (1963) notes that ‘the ratio of two CMFs is a CMF but the ratio of two SMRs is not an SMR’.

In spite of these criticisms, the SMR and the CMF usually provide numerical results that are remarkably close in practice. In cases in which they differ, moreover, it is not necessarily true that the CMF is more nearly ‘correct’. Table 2.14, compiled by the Office of Population Censuses and Surveys (1978), examines occupations of British men for which the CMF and SMR differ by 5 or more when each was expressed as a percentage. The large discrepancy between the two measures for trainee craftsmen in engineering trades is caused by undue weight in the CMF to the lack of deaths among men over 25 years of age, even though these men accounted for only 1% of the population. Since only one death in the 55–64-year age group would have increased the CMF from 1 to 35, this is another example of its extreme sensitivity to small numbers and serves as a reminder that we need to consider variance as well as bias in our choice of a summary statistic (see also Example 2.6).

The CMF may itself overemphasize biases in the basic data. As noted by the Office

Table 2.14 Occupation units for which the CMF and SMR differ<sup>a</sup>

Occupation unit		SMR (%)	CMF (%)		Age group (years)				
No.	Title				15-24	25-34	35-44	45-54	55-64
Major differences (10 or more)									
009	Workers below ground	112	138	Mortality ratio	139	112	61	46	184
				% Population	14	33	26	19	8
032	Trainee craftsmen (engineering trades)	26	1	Mortality ratio	27	0	0	0	0
				% Population	99	1	0	0	0
117	Pilots, navigators and flight engineers	86	151	Mortality ratio	190	82	26	80	195
				% Population	11	23	33	31	2
151	Fire brigade officers and men	99	117	Mortality ratio	102	83	63	66	145
				% Population	15	33	23	21	8
152	Police officers and men	109	213	Mortality ratio	76	68	66	72	299
				% Population	22	32	27	17	2
187	Chiropodists	74	101	Mortality ratio	805	145	94	79	67
				% Population	4	9	18	30	39
302	Metallurgists	96	107	Mortality ratio	—	85	64	114	117
				% Population	27	37	18	11	6
221	Armed forces (UK)	147	252	Mortality ratio	123	100	100	118	337
				% Population	43	31	17	7	1
222	Armed forces (foreign)	149	243	Mortality ratio	198	128	83	129	314
				% Population	41	33	21	4	1
Minor differences (5-9)									
61	Shoemakers and repairers	156	150	Mortality ratio	47	127	154	132	164
				% Population	10	14	20	23	33
114	Other labourers	201	209	Mortality ratio	292	369	285	113	81
				% Popoulation	23	15	16	20	26
134	Lorry drivers' mates van guards	120	128	Mortality ratio	97	118	115	153	123
				% Population	69	12	9	6	5
158	Domestic housekeepers	117	125	Mortality ratio	—	—	241	210	94
				% Population	—	5	16	29	50
163	Kitchen hands	113	123	Mortality ratio	172	253	225	147	91
				% Population	21	12	16	23	28
164	Maids, valets and related workers	121	129	Mortality ratio	198	200	190	151	105
				% Population	15	19	17	20	30
204	Chemists	89	95	Mortality ratio	104	53	66	96	100
				% Population	14	39	23	14	9
All men		100	100	Mortality ratio	100	100	100	100	100
				Population (15 - 64 = 100)	23.5	20.1	18.9	19.5	18.1

<sup>a</sup> From Office of Population Censuses & Surveys (1978)

of Population Censuses and Surveys for the data in Table 2.14:

'Pilots, policemen, firemen and members of the armed forces all recorded differences between the SMR and CMF of greater than 10. Although men in these units generally retired before 55 years of age and took up other work their main occupations instead of their last occupations were often recorded when the deaths were registered. The age-specific mortality ratio in the 55-64 age group was consequently inflated by this

bias. The CMF placed considerable weight on the death rates in this age group since over 60 percent of the standard deaths occurred at this age. Since however for each of these occupation units the population in this age group was small, the SMR placed less weight on these high death rates, compensating for the bias introduced.'

Silcock (1959) determined analytically three conditions under which the CMF and SMR give substantially different results. Denote by  $p_j = n_j/N$  and  $p_j^* = n_j^*/N^*$  the age distributions of the comparison group and the standard population, both expressed as proportions, and by  $\hat{\lambda}_j$  and  $\lambda_j^*$  the corresponding rates. Then the conditions are: (i) the differences  $p_j - p_j^*$  must be non-negligible; (ii) the ratios  $r_j = \hat{\lambda}_j/\lambda_j^*$  must vary substantially with age ( $j$ ); and (iii) the differences in (i) and the ratios in (ii) must be correlated, such that positive differences tend to occur with high ratios and negative differences with low ratios, or *vice versa*. The 'data' in Table 2.13 confirm that these conditions hold in situations where the CMF and SMR differ.

(g) *Summary ratios under heterogeneity of effect*

Much of the preceding discussion of the relative merits of the CMF and SMR was conducted under the (sometimes only implicit) hypothesis that each measure was estimating the same quantity, namely the ratio of age-specific rates assumed constant from one age group to the next. The major exception was the fact just cited that the two measures could yield substantially different results only if the age-specific ratios varied in tandem with differences in the age distribution. Our basic viewpoint remains that summary measures should be avoided whenever there is substantial heterogeneity in the age-specific quantities being summarized.

Other authors have been more concerned with the issue of how to choose a summary measure in order to arrive at a scientifically meaningful result, even in the face of heterogeneity. Greenland (1982) notes that the CMF, viewed as a weighted average of ratios with weights equal to standard deaths (equation 2.5), represents the proportionate increase (or decrease) in the total disease rate that would be expected to occur in the standard population if its members had the same exposures as those in the cohort. Similarly, the SMR represents the proportionate increase in the cohort disease rate due to the exposures that occurred as a result of cohort membership. Following Miettinen (1976), he proposes yet another summary measure that uses the age distribution of the combined (cohort plus standard) population for calculation of the weights used to multiply the age-specific ratios, namely:  $u_j = (n_j + n_j^*)\lambda_j^*$ .

Several other proposals are reviewed by Kilpatrick (1963) and Fleiss (1973). The relative mortality index (RMI) weights the ratios  $r_j = \hat{\lambda}_j/\lambda_j^*$  by the age distribution of the cohort:

$$\text{RMI} = \frac{\sum_{j=1}^J n_j \hat{\lambda}_j / \lambda_j^*}{\sum_{j=1}^J n_j} = \frac{\sum_{j=1}^J d_j / \lambda_j^*}{\sum_{j=1}^J n_j}.$$

Liddell (1960) examines some properties of this measure, which he credits to Kerridge (1958) and Doering and Forbes (1939). Yerushalmy (1951) and Elveback (1966) use the length of the age interval to weight these same ratios, while Haenszel (1950) considers a ratio of directly standardized rates with weights equal to the number of

years of working life (up to age 65) lost. Of course, if the ratios are constant, none of these schemes is optimal in the sense of minimum variance, as was true of the SMR, nor has any of them come into common use.

## 2.4 Proportional measures of incidence and mortality

Proportional mortality studies were mentioned in §1.6 as a timely and cost-effective way to provide a rough measure of the effect of cohort membership or of specific exposures on mortality caused by particular diseases. If the only data available concern incident cancer cases, information about their distribution by topographic site or histology can alert one to the possibility of unusual patterns of incidence that should be investigated using more orthodox methodologies. It is especially important in such studies to try to ascertain all deaths or cases that occur in a defined population during a defined period, or else to ensure that the probabilities of ascertainment do not depend on the cause of death or type of cancer. Otherwise one runs the risk of making misleading inferences due to selection bias. Even if such precautions are taken, major problems of interpretation remain due to the logical impossibility of making comparative statements about rates from 'numerator' data only.

In this section we present the usual epidemiological methods for adjusting numerator or proportional mortality data so as to account for the differences in age distribution between the study group and the standard population. The same techniques may be used to control the effects of calendar year and other potentially confounding variables. We also show some empirical comparisons of the different results obtained from SMR and proportional mortality analyses when both are applied to the same set of data. A more theoretical evaluation of the behaviour of the proportional mortality measures is presented in §4.7, together with some suggestions for statistical modelling of this type of data.

### (a) *The proportional mortality ratio (PMR)*

The basic idea of proportional mortality analysis is to compare the fraction of cohort deaths due to a specified cause with the corresponding fraction for the general population. Denote by  $d_j$  the number of deaths from the cause of interest observed in age interval  $j$  in the study group, and by  $d_j^*$  the corresponding number of standard deaths. Likewise denote by  $t_j$  and  $t_j^*$  the total numbers of age-specific deaths, regardless of cause. Then, with  $D = \sum d_j$ ,  $D^* = \sum d_j^*$ ,  $T = \sum t_j$  and  $T^* = \sum t_j^*$  indicating the totals of these quantities summed across age strata, the ratio  $(D/T) \div (D^*/T^*)$  of the two proportions provides a crude measure of relative effect. Note that this can also be expressed as the ratio of the observed number of cause-specific deaths,  $D$ , to the number expected by applying the standard proportion to the total deaths, namely  $T \times (D^*/T^*)$ .

Since the death rates, and thus the proportions of deaths for different causes, depend on age in different ways, the age distribution of the comparison group can influence the overall proportion  $D/T$ . The conventional approach to adjusting for such age differences (Monson, 1980) is to calculate the expected numbers on an age-specific

basis and then sum up. This yields a measure known as the proportional mortality ratio (PMR), namely:

$$\text{PMR} = \frac{D}{\sum_{j=1}^J t_j(d_j^*/t_j^*)} \quad (2.16)$$

Under the null hypothesis that the age-specific proportions agree between cohort and standard population, the PMR will be approximately unity in large samples of data, regardless of differences in the age distributions. However, unless the disease is rare, the PMR does not estimate any well-defined or appropriate parameter under alternative hypotheses of interest (§4.7). For this reason, we do not recommend that statistical inference procedures be conducted on the PMR, but suggest instead that one use the parameter estimates, tests and confidence limits produced as a by-product of the model fitting described later. Nevertheless, we record here an equation for the standard error of  $\log(\text{PMR})$  that is based on the notion that the denominator is fixed whereas the numerator  $D$  is a sum of independent binomial variables:

$$\text{SE}(\log \text{PMR}) = \frac{(\sum_{j=1}^J d_j(t_j - d_j)/t_j)^{1/2}}{D} \quad (2.17)$$

The simpler equation (2.9) can be viewed as a conservative approximation to (2.17), to be used with the PMR as well as the SMR provided that the fraction of deaths due to the cause of interest is quite small.

#### (b) The PMR and the relative SMR

Several investigators have noted that, in practice, the PMR for a given cause of death is approximately equal to the SMR for that cause divided by the SMR for all causes combined. When there is no stratification by age or other factors one has:

$$\text{PMR} = \frac{D/T}{D^*/T^*} = \frac{D/N}{D^*/N^*} \div \frac{T/N}{T^*/N^*} = \frac{\text{SMR}}{\text{SMR}(\text{all})}, \quad (2.18)$$

where  $\text{SMR}(\text{all})$  denotes the all-causes ratio (Decouflé *et al.*, 1980). This equality does not hold, however, for the usual age-standardized PMRs and SMRs. Kupper *et al.* (1978), who refer to the ratio of cause-specific to all-causes SMR as the 'relative standardized mortality ratio' (RSMR), have attempted to establish confidence bounds within which the ratio of the two sides of the equation would be expected to lie with high probability. Unfortunately, their method relies on an assumption that cannot be verified from numerator data alone, namely that the age distribution of the deaths that would be expected to occur in the comparison group by applying the standard rates is roughly the same as the age distribution of the standard deaths. If this condition does not hold, the magnitude of the difference between the two sides of the equation could be larger than their calculations would suggest. Nevertheless, it is a frequent empirical finding that the PMR and RSMR tend to agree, probably because the age distributions in question are rarely all that different.

Table 2.15, adapted from Decouflé *et al.* (1980), illustrates the typical agreement

Table 2.15 Cause-specific mortality experience of non-white male foundry workers employed at a plant between 1938 and 1967<sup>a</sup>

Cause (ICD, 7th revision)	No. of observed deaths	SMR (%)	PMR (%)	RSMR (%)
All causes	172	55	100	100
Cancer (140–205)	35	88	153	160
Digestive (150–159)	14	92	161	167
Respiratory (160–165)	12	114	199	207
All other	9	65	111	118
Stroke (331, 332)	9	36	62	65
Circulatory disease (400–468)	54	53	93	96
Accidents, suicide, homicide	27	52	107	95
All other causes	47	48	92	87

<sup>a</sup> Adapted from Decouflé *et al.* (1980)

SMR = standardized mortality ratio

PMR = proportional mortality ratio

RSMR = relative standardized mortality ratio

found between PMRs and RSMRs for the same cause. In their example, the overall mortality of the workers was so low in comparison with that of the general population that there appeared to be a marked excess of respiratory cancers when numerator data alone were considered. If one believed that the selection bias that operates to make industrial workers healthier than the general population applied with equal force to all diseases, then it would be reasonable to conclude that the elevated PMR observed for respiratory cancer was indicative of an effect of exposure on that disease. However, a more plausible explanation in this case is that the elevated PMR results from the selection bias being less pronounced for cancer than it is for other diseases (Enterline, 1975). In view of the uncertainty surrounding these assumptions, use of the RSMR and PMR remains controversial (Wen *et al.*, 1983).

Our last example illustrates a number of the basic calculations introduced throughout the chapter using data from the Montana cohort.

#### Example 2.9

Table 2.16 presents the CMFs, SMRs, PMRs and RSMRs for four causes of death for the 8014 workers in the Montana cohort. The 1950 US standard population (Table 2.5) provided the weights used for direct standardization and calculation of the CMF. Expected numbers needed for determination of the SMRs were obtained by multiplying the exact person-years shown in each cell of Table 2.2 by the corresponding rates for US white males (Appendix III) and then summing. The denominators of the CMF statistics were obtained by applying the standard weights, a function of age alone, to the standard rates, which varied by both age and year. Thus, we have used equation (2.3) to calculate the CMF, rather than the simpler equation (2.4), which applies only when the weights are proportional to the denominators of the standard rates. The PMR was determined from equation (2.16); and the standard errors for the logarithms of the CMF, SMR and PMR (which equal the standard errors of the estimate expressed as a percentage of the estimate) were determined from equations (2.7), (2.9) and (2.17), respectively.

This working cohort was unusual in having an all-causes summary rate ratio (CMF or SMR) substantially above 100%. The PMRs and RSMRs show good agreement, as do the CMFs and SMRs. Note that the standard errors of  $\log(\text{PMR})$  are less than those for  $\log(\text{SMR})$ . Part of the difference is due to the inherently smaller degree of variability in a proportion than in a rate. For example, had we used equation (2.9) to find the standard error of  $\log(\text{PMR})$  for circulatory diseases, the result would have been  $\text{SE}(\log \text{PMR}) = 2.6\%$  rather than 1.9%. Similarly, the  $\log(\text{SMRs})$  have smaller standard errors than do the  $\log(\text{CMFs})$ . Besides



Table 2.16 Mortality ratios for the Montana cohort

	Cause of death			
	All causes	All cancers	Respiratory cancer	Circulatory disease
No. of observed deaths	3404	621	288	1535
No. of expected deaths	2761.0	485.4	137.1	1473.4
CMF( $\times 100$ ) $\pm$ SE(%) <sup>a</sup>	111.4 $\pm$ 2.5%	127.8 $\pm$ 5.0%	234.6 $\pm$ 7.0%	93.1 $\pm$ 3.1%
SMR( $\times 100$ ) $\pm$ SE(%) <sup>a</sup>	123.3 $\pm$ 1.7%	127.9 $\pm$ 4.0%	210.1 $\pm$ 5.9%	104.2 $\pm$ 2.6%
PMR(%) $\pm$ SE	100	102.9 $\pm$ 3.6%	166.4 $\pm$ 5.5%	84.5 $\pm$ 1.9%
RSMR(%)	100	103.7	170.4	84.5

<sup>a</sup> Standard errors are computed on a log scale, e.g.,  $SE(\log CMF) = 0.025$ , and we thus express the standard deviations of the estimate as a percentage of the estimated value

CMF = comparative mortality figure

SMR = standardized mortality ratio

PMR = proportional mortality ratio

RSMR = relative standardized mortality ratio

the excesses of respiratory cancer and circulatory disease (mostly diseases of the heart) shown in Table 2.15, there were more deaths observed from tuberculosis, cirrhosis of the liver and emphysema than would have been expected from the standard rates (Lee-Feldstein, 1983).