J Clin Epidemiol Vol. 41, No. 12, pp. 1217–1237, 1988
Printed in Great Britain.

0895-4356/88 $3.00 + 0.00

Special Communication

THE DEVELOPMENT OF COHORT STUDIES IN EPIDEMIOLOGY: A REVIEW

F. D. K. LIDDELL

Department of Epidemiology and Biostatistics, McGill University, 1020 Pine Avenue West, Montreal, Quebec, Canada H3A 1A2

(*Received in revised form* 30 *June* 1988)

Abstract—An historical outline of the evolution of cohort (or incidence) studies spans well over 100 years, from the work of Farr and Snow in the 1850s, through an appraisal of analytical methods in 1977, after which the literature mushroomed. Since the early 1950s, analysis has conventionally taken the form of comparing subcohorts that had suffered varying degrees of exposure to factor(s) under investigation. For this purpose, the "subject-years" approach to data reduction has now become virtually universal. Usually, some population's mortality (or morbidity) experience is used as reference, but there is continuing controversy over the choice of reference population, while difficulties arise in relation to study intervals, periods over which exposures should be measured, etc. The material for analysis becomes age- and periodic-specific ratios of disease, which, collapsed over ages and periods, lead to Standardized Mortality (or Morbidity) Ratios. For the analysis itself, Poisson regression models are efficient. From the late 1970s, analysis by case–referent methods has become common; here, the debate centres on how closely, and in what ways, referents should be matched with the cases. Logistic regression is the most common form of analysis. As there have been excellent recent summaries of methods of analysis (for both approaches), little emphasis is placed here on those aspects of development. Comparisons are made of research designs, and some possibilities for future development are outlined.

Age–period–cohort analysis Case–cohort Case–minicohort Case–referent-within-a-cohort Cohort analysis Exposure measures "Healthy worker effect" Incidence studies Logistic regression Occupational cohorts Reference population Study intervals Subject-years Transient states

1. INTRODUCTION

1.1. Definitions; scope of the review

In an important chapter, Doll [1] explained that the prospective (or cohort) method of inquiry, consisting in "the delineation of a group of persons who are distinguished in some specific way from the majority of the population and observation of them for long enough to allow any unusual morbidity or mortality to be recognized", is extremely simple in concept. So simple indeed that most epidemiologists generally regard a "cohort" as any defined group of people who are followed forward, individually, from the time at which each first suffered some common demarcating circumstance, or "exposure". Thus, the current **Dictionary of Epidemiology** [2] lists five phrases merely as "synonyms" for cohort studies—although they fall into at least four classes, depending *inter alia* on: whether the members of the cohort are presumed healthy or diseased at the onset of the exposure; whether or not such onset is at the discretion of the investigator; the nature of the exposure, e.g. therapeutic agent, environmental (including occupational), or other; and what type of outcome in the follow-up period (e.g. incident or serial change) is of interest.

Cohort studies are usually of non-experimental, or "observational", design. However, *randomized trials* (including the RCTs that Cochrane [3] considered so important) are cohort studies in which the exposure (or therapeutic agent) has been assigned with an

experimental plan. What may be termed *clinical cohorts* consist of people known to be diseased, who are followed to determine the (clinical) outcome. These studies are often, but not exclusively, of few patients followed for short periods of time. Another class could be designated *repeated-measures cohorts*; these are studies of serial change in the values of (usually) several variables, such as pulmonary function tests, presumed indicators of ischemic heart disease, or the "normal" physical development of children. It is characteristic of this type of cohort that the main outcome is not an incident. In *epidemiologic (or public-health) cohorts*, the population followed is usually assumed healthy when the cohort is assembled (i.e. before the onset of the exposure of interest) and is commonly followed to examine the effects (usually in terms of incidence) of exposure to one or more environmental agents, such as those of daily life—e.g. cigarettes or diet—or of occupational origin (not to a pharmaceutical agent nor to medical treatment). [There is also an essentially different class of study for which the word "cohort" was originally proposed, as will be explained in Section 1.4.]

No review could cover adequately all classes even of observational cohorts, because each class has different methods of assembling and following subjects, different ways of making measurements on the subjects and on their exposures, and entirely different forms of analysis—to achieve essentially different aims. Although the emphasis in this review, following Doll [1], is on epidemiologic cohorts, it is necessary to mention—particularly in the historical outline of Section 2—certain clinical cohorts (which I call "follow-up studies", using the terminology of Truelove [4]) and repeated-measures cohorts (termed "longitudinal studies", following common biostatistical usage). The interplay between the analysis of results from clinical trials and cohort studies will also be brought out where appropriate.

The present review is then of the development of "closed" cohorts (see Section 1.4) in which the outcome is—at least for the purposes of the specific study—unrepeatable; examples include congenital defect, the first signs of progressive disease, and, perhaps most commonly, death. Because of my own epidemiologic experience, extra emphasis is placed in this review on occupational cohort studies; fortunately, most of the recent advances have been made in this field, and the principles apply to all cohorts of the "public health" genre.

1.2. Exposure–response relationships

The importance of demonstrating an exposure–response relationship (whereby the more severe the exposure the greater the incidence of disease) in attempting to establish causation (rather than mere association) [5] cannot be over-emphasized. Doll fully recognized this when he made the point that it is desirable *if at all possible* to divide the group (or cohort) into several "subcohorts" with varying degrees of exposure to the factor(s) under investigation; ideally, one subcohort should consist of subjects who have had no unusual exposure at all [1]. The experience of the various subcohorts may then be related quantitatively to the degree of exposure. (The 1956 report on mortality among British doctors in relation to smoking habit [6] is an important early example; see Section 2.4.)

From the 1960s, comparisons between subcohorts formed the *raison d'être* of most modern cohort studies; see Section 2.4. In this approach the methodology of Sections 3 and 4 was usually adopted. However, there were inevitably situations where it was impossible; one reason is that all members of the cohort had been exposed to a similar degree. In the last decade, the elucidation of exposure–response relationships has not only maintained its importance, but has also been made more readily achievable; see Sections 5 and 8.3.

1.3. Terminology

To describe the method of inquiry reviewed here, Doll [1] used mainly the term "prospective study", allowing "cohort study" as a synonym; "incidence study" would also appear appropriate [2]. A study which starts with a population of known characteristics and works *forward* to the sick persons in that population has long been described in epidemiology as "prospective", and studies which work *back* from the sick persons to the factors which may have led to their illness as "retrospective". These words do not place an enquiry in time but—because they can cause misunderstanding among those who expect them to have their everyday meanings [7]—they have been largely replaced by "cohort" and "case–control", despite the drawbacks of these terms. [Although Frost [8] used the word "retrospective", this was to describe the historical family records he had used, not to

characterize what was undoubtedly a cohort study.]

1.4. Age–period–cohort analysis; hypothetical and "closed" cohorts

Objections to the use of "cohort" (in the sense of Section 1.3) were raised in 1979 by Springett [9] and Jacobs [10] because it had already been used by Wade Hampton Frost [11]. Indeed it had, but rather differently—for analysis of mortality, published in cross-sectional form, but exploiting the facts that persons who were, for example, aged 50–59 years in the 1930 Vital Statistics, were aged 0–9 in 1880, 10–19 in 1890, etc. According to Frost [12] in 1939, this "cohort analysis" had been used as early as 1930 by K. F. Andvord (in a Norwegian journal). Its first use for non-infectious disease was in 1951 by Korteweg [13], and it has often been exploited for infectious diseases, as for example in 1954 by McDonald and Springett [14]. A simple form of this analysis is now a standard statistical technique [15].

Perhaps unfortunately, the analysis of a prospective study is often given the same name [16, 17], so the earlier technique has now been called "age–period–cohort analysis" [17–19]. Here, the cohort is—obviously—merely hypothetical, and thus clearly distinct from the "closed" cohort of our present concern; this Case and Lea [20] stated to be "one in which the ... life experience of each of its members, subsequent to the entry of that member into the observational field, is of relevance to the analysis, regardless of whether or not the member has ceased to belong to the particular defined class ... used in circumscribing the observational field when designing the survey".

1.5. "Typical" and "atypical" cohorts

In an illuminating exposition, Breslow [17] remarked that: "Epidemiologic cohort studies *typically* involve the follow-up of *large* population groups *over many years* to ascertain the effects of environmental exposures on the outbreak of illness and the age and cause of death" (all emphasis added). Nevertheless, by no means all epidemiologic cohort studies have these typical features. For instance, in 1958, Hill *et al.* [21] reported on the first four cohort studies, in three countries, concerned with the risks of major infant malformations following maternal rubella during pregnancy, and Doll [1] cited them as exemplars of cohort studies: yet the four cohorts comprised, in all, only 104 mothers, while follow-up was for little longer than 9 months. There have been many other examples; see Section 2.

The term "atypical" is used below to describe a cohort study, of the class under review, in which either the population group was comparatively small or follow-up was for a short period of time.

1.6. Modes of reasoning

It will have been noted that the comparison of subcohorts with varying degrees of exposure, postulated [1] as so desirable (see Section 1.2) implies reasoning from cause to effect. However, the other mode—from effect to cause—has been exploited, particularly in the last decade.

The earlier, or conventional, method requires the delineation of subcohorts, each defined in terms of exposure to the variable(s) of interest, followed by the reduction of the truly massive volume of data to a few simple measures of mortality for each subcohort (see Section 3), and thereafter comparison of the subcohorts, one with another, in terms of these mortality measures (see Section 4).

In more recent approaches, cases are identified after follow-up, and referents are selected in one of several possible ways; cf. Section 8.3. Thereafter the analysis follows well-established routines (described by *inter alia* Breslow and Day [22]) for case–referent studies. An essential feature of this "case–referent-within-a-cohort" approach is that follow-up has to be completed before all the cases can be determined. Thus, there is a radical difference from the traditional case–control (or retrospective) study [1], in which cases are determined as the first stage of the research, and referents are selected in a second stage—both independent of any cohort study. Nevertheless, cohort studies can be designed in which the only analysis is to be by case-referent methods.

2. HISTORICAL OUTLINE

2.1. The time element in the 19th Century

This outline starts with two important 19th Century studies, in which the time element was essential. The "statistical method" was in use before 1835, when Pierre Charles Alexandre Louis, of the medical faculty in Paris, reported [a second time (at the Publisher's request)] a clinical investigation (completed some years

earlier) which helped end the vogue of bleeding as a panacea [23, 24]. A total of 77 patients with pneumonia had been bled at different intervals after the onset of illness; a slightly free translation follows: "When all patients were classified without regard to outcome but according to when they were bled, we have:- The total first bled within 4 days was 41, of whom 18 (or 3/7) died. The total first bled in days 5–9 was 36, of whom 9 (or only 1/4) died." [This finding would not nowadays be considered of statistical significance.] During 1844–48, Ignaz Semmelweis sought to determine the causes of puerperal fever [25]. After he had required all medical students to wash their hands in a solution of chlorinated lime before examining any woman in labour, the mortality from childbed fever in the following year fell to 1.3% (from around 8%). Later, following the examination of a woman in labour suffering from a "foully discharging carcinoma of the uterus", he and his associates "merely washed [their] hands with soap", and then examined 12 women, 11 of whom died of puerperal fever—support for a hypothesis obtained by ingenious, if drastic, means.

In October 1853, William Farr had persuaded the Registrar General to ask each London Water Company to indicate the source(s) from which water had been obtained since 1849 for the supply of London districts, and to provide several other details. Farr had reported that the risk of death from cholera in the epidemics of 1848–49 and 1853 depended clearly on the water source, as well as on the elevation of the district [26, 27]. John Snow seized on the 1854 epidemic in the Kennington area as a "Grand Experiment", because the pipes of the Southwark & Vauxhall Water Company (SVC)—supplying "the most impure water"—and those of the Lambeth Water Company (LC)—supplying "purer, filtered water"—were "intimately mixed". Snow identified the water supply to each house in which any person dying from cholera had resided, and found many times more deaths in houses within the "experimental" area supplied by SVC than in those supplied by LC. However, the denominators (although perhaps roughly equal) could not be ascertained accurately. Nevertheless, during the 14 epidemic weeks of 1854, in the 40,046 houses (in total) supplied by SVC, cholera mortality (4093 deaths) was over 5 times higher than in the 26,107 houses supplied by LC (461 deaths) [27, 28].

The full findings formed an invaluable link in the chain of evidence being forged by Snow [28]. (However, just what impurity of the water had led to the high risk of cholera remained for many years a matter of controversy: Snow's explanation—correct, but far ahead of its time—of the role of infection was rejected by the Anticontagionist Movement; see also Terris [29].)

This investigation was of unusual design, termed by the present author "case–cohort" (see also Section 8.3). The numbers of houses in which the cohort resided were known according to source of water supply, and the relevant populations could be estimated, in advance, to form the denominators of incidence rates, but the persons forming the cohort were not followed individually during the epidemic. Rather, when a cholera death was notified, the source of water supply to the house in which the deceased had resided was identified, so that numerators could be built up. As less than 4% of all houses were involved, the design was highly efficient.

2.2. Goldberger on pellagra

Terris [30] considered what Wade Hampton Frost had called Snow's "nearly perfect model" of epidemiologic investigation (which "remained necessarily at the level of observation") rather less complete than Goldberger's work on pellagra (which could "achieve the additional certainty of experimental demonstration"). Of the 17 papers Terris chose "to represent the most significant publications on pellagra by Joseph Goldberger and his colleagues", only four (all pubished in 1920) are relevant here. They described one complex survey of pellagra incidence in seven cotton-mill villages of South Carolina in 1916, in relation to: diet; sex, age, season, occupation, and "disabling sickness"; factors of a sanitary character; family income and other economic factors. There were horrendous problems, not least that there were only 115 definite incident cases among a cohort of 4160 persons living in 750 households. This meant that numbers in subcohorts defined by levels of any one factor were small; the search for interactions was painstaking, but could hardly be powerful.

Two typical tables have been condensed (correcting a few minor errors in rates) for this review. Table 1 shows the remarkable differences in pellagra incidence by age and sex. Table 2 (based on only 56 households), suggests an important protective effect of milk (the risk

Table 1. Pellagra incidence by sex and age in seven cotton-mill villages of South Carolina during 1916 (source: Ref. [30] p. 201)

	Population		Cases†		Incidence§	
Ages* (yr)	M	F	M	F	M	F
Under 2	122	119	0	0	0.0	0.0
2–4	217	190	3	13	13.8	68.4
5–9	313	296	20	15	63.9	50.7
10–19	521	537	7	7	13.4	13.0
20–54	817	849	5	39	6.1	45.9
55 and over	81	98	5	1	61.7	10.2
All ages	2071	2089	40	75	19.3	35.9

*"Classified into such periods as indicate the greatest variations".
†Definite cases of pellagra in the white population.
§Cases per thousand population.

of a household becoming pellagrous varied nearly 5-fold according to the recorded milk supply), but without information on the distributions by age and sex within the three groupings of households.

2.3. *Early uses of life-table methods*

In his seminal paper of 1933, Frost [8] described with great clarity a study of the risks of persons in familial contact with pulmonary tuberculosis. The life-experience and mortality of all 794 present and former members of 132 negro families constituting practically the entire negro population of Kingsport, Tenn., were summarized, for each year of age from birth to over 60. The main finding was that the annual (age-adjusted) rate of death from tuberculosis, per 1000, was 4.6 ± 1.01 for the 299 with positive contact history, but only 2.3 ± 0.38 for the "negatives". Frost [8] stated that the method (see Section 3.1) was in no sense new in its application to studies of tuberculosis, having been used in 1910 by William Palin Elderton and S. J. Perry; Frost's claim to novelty was in the use of historical records.

2.4. *Developments since World War II*

Mention must be made of the "1946 cohort" in Britain [31]. Initiated with short-term aims, namely to answer questions on the availability, use, effectiveness, etc. of the maternity services, to different social classes and in different parts of the country, it concerned 13,687 women (over 80% of the total) who gave birth in Great Britain during a single week in March 1946. The opportunity was "too good to miss" and Atkins *et al.* [32] listed 27 longitudinal studies completed by 1981 of a random sample (with some imperfections) of 5362 mothers and/or their children or even grandchildren: a "sperm to worm" concept keeps the children born in 1946 under review, but a mortality study with sufficient deaths to be worthwhile cannot be developed yet. It is likely that the "1946 cohort" will long remain a longitudinal study, with multitudinous outcomes, mainly changes over time (between each occasion on which measurements are made) in a highly complex multivariate set of observations. The statistical analyses required are quite different from those for cohort studies germane to this review, as is made patent in the one statistical chapter [33] of Mednick and Baert's **Prospective Longitudinal Research** [34]. Further "cohorts" were set up in the "same" week in 1958 and again in 1970. Special mention should be made of the longitudinal studies—called the National Child Development Study—of the 17,414 babies born in 1958, of whom over 70% (i.e. 12,537) were still in view 23 years later [35]; the history and aims of this study, the methods used and the data sources available were reviewed in 1985 by Shepherd [36].

The first two occupational cohort studies were reported in 1952 and 1955 by Doll [37, 38], who used for his analysis a close approximation (required in the absence of modern computers) to the "subject-years" method (see Section 3). The second study—of lung cancer in 113 asbestos workers who had been employed for at least 20 years in areas scheduled under Regulations as being dusty—is the more instructive, although atypically small. Doll's conclusion was

Table 2. Pellegra incidence by household milk supply during 1916 among households of cotton-mill workers in seven villages of South Carolina (Source: Ref. [30] p. 165)

	Households			
Household milk supply*	Total	Number affected with pellagra	Per cent affected	"Relative Risk"
7.0 and over	311	12	3.9	1.0
1.0–6.9	262	16	6.1	1.6
Less than 1.0	154	28	18.2	4.7
All amounts	727	56	7.7	—

*Quarts per "adult male unit", for a 15-day period between 16 April and 15 June, 1916.

Table 3. Summary of findings by Case and Lea [20]. Deaths found and, in parentheses, ratio of deaths found to deaths expected

Mortality, 1930–1952	Cohort (i) Mustard gas	Cohort (ii) Bronchitis	Cohort (iii) Amputation
All causes	547 (1.5)	932 (1.4)	383 (1.0)
Cancer of lung and pleura	29 (2.1)	29 (2.0)	13 (0.8)
Other neoplasms	50 (1.1)	75 (0.9)	59 (1.0)

that "lung cancer was a specific industrial hazard of certain asbestos workers": the number of deaths expected was no higher than 1.1, whereas 11 had been observed [38] (see also Section 3.6).

Case and Lea [20] described in 1955 an investigation of three cohorts of men receiving compensation at the start of 1930 for (i) mustard gas poisoning (1267 men), (ii) chronic bronchitis (1421 men) and (iii) single-leg amputation (1114 men), occurring as a result of the 1914–18 war, followed (without loss) to end-1952. The findings, in Table 3, indicated, for cohorts (i) and (ii), marked excesses of total mortality and of deaths from cancer of lung and pleura, but not from other neoplasms—interpretation is a matter of controversy, not addressed here. However, cohort (iii)—who had not suffered any unusual life-threatening exposure after 1930—had mortality close to that of all males in England and Wales.

In each of the three cohorts of Case and Lea [20], the exposure had been essentially "one-shot". Perhaps the most important cohort involving one-shot exposures was of the survivors of the atomic bomb catastrophes in 1945, a great strength lying in the fact that the exposures were not of one intensity; a recent description is in Ref. [39]. One-shot exposures result from disasters or in industrial settings which existed for only a short spell relative to a person's life-time. Examples of this type have been reported on workers assembling gas masks (for the armed services) containing crocidolite filters or manufacturing the filters: 951 females in Nottingham, England [40], and 199 workers, men and women, in three factories of Eastern Canada [41]. Exposures were known to be of very limited duration, but such proliferations of deaths from malignant mesothelial tumours, occurring 18 years or more after first exposure, have cast much light on the aetiology of this rapidly fatal, but fortunately rare, tumour.

In 1956, Doll and Hill [6] reported on lung cancer and other causes of death in relation to smoking in a cohort of 34,000 male British doctors. Briefly, about two-thirds of all members of the medical profession in the U.K. responded to a questionary about their smoking habits, distributed at end-October 1951. Four subcohorts of males aged 35 years and above were formed: non-smokers (roughly 17% of the cohort) and "men smoking a daily average of" the following amounts: 1–14 g (34%), 15–24 g (31%), 25 g or more (18%); and found annual (age-adjusted) death rates from lung cancer, per 1000, as 0.07, 0.47, 0.86, and 1.66, respectively. The methods of data-reduction and analysis were especially interesting (see Section 3.4 and Appendix B). The epidemiological findings have been fully discussed [42], but it must be emphasized here that the demonstration of the emphatic exposure–response relationship was crucial to the argument that cigarette smoking was causally related to lung cancer (disqualifying other theories such as genetic susceptibility) [35].

The first analysis based on exposure histories for only samples of non-cases (but for all cases), was initiated by Court Brown and Doll in 1955 and published by the (British) Medical Research Council (MRC), in a classic report, 2 years later [43]; the mode of argument, following the convention of the day, was from cause to effect. (Doll [44] comments: "The conduct of the work was a fascinating experience as we were given precisely 6 months to obtain a result and assured that any help required would be provided by the MRC. In the event 100 people participated in the work and we provided results 6 months to the day. These however were criticized by one member of the MRC's Radiology Committee who thought—correctly—that we had been given the wrong advice about how to assess radiation dose, and we had the best part of another year's work to get the results published in the MRC report.")

McDonald [45] herself interviewed 93% of 3295 pregnant women in England in 1952–55; the outcome was determined for 3216 (3259 fetuses). Innovatively, the analysis was from effect to cause: for example, of the 50 mothers of infants with major defects, 10 (i.e. 20.0%) had been engaged on heavy manual work

during the first 12 weeks of pregnancy, but only 8.0% of the mothers of 3059 normal infants. (Part of the then conventional analysis—from cause to effect—is given by Doll [1].) Meanwhile, the "atypical" cohort studies [21] mentioned in Section 1.3, of the effects of rubella during pregnancy, are important to the historical flow.

From the continuing longitudinal study based on the population in 1949 of Framingham, Mass. [46], Jerome Cornfield created an incidence study: he identified from the original sample (1329 males aged 40–59, clinically examined between 1948 and 1952) 92 members who, during the 6-year period following the initial examination, developed clinically manifest coronary heart disease (myocardial infarction or angina pectoris). In 1962, he explained, in a characteristically powerful paper [47], how the 92 attacked could be differentiated from the remaining 1237 by means of discriminant analysis, in the form now known as logistic regression analysis. He then examined the joint dependence of serum cholesterol and blood pressure as risk factors for coronary heart disease. It should be noted that the mode of reasoning was from effect to cause, paving the way for the change in direction of analysis that became common in the 1980s. However, the point of attack (within the 6-year follow-up period) was disregarded. Justification for random sampling of all non-cases, to reduce the work of estimating exposures, was given in 1973 by Mantel [48] and extended in 1975 by Kupper *et al.* [49].

In 1965, Enterline [50] improved on earlier approaches to data reduction and analysis, and 8 years later Enterline and Henderson [51] stated that this method was then widely used. They referred to it as the "modified life-table" method (which might more reasonably be a name for the approach of Case and Lea [20]), but it is now usually known as the "subject-years" (or "person-years", or "man-years") method [52]; see Section 3 below.

Higgins and Keller [53] reported in 1970 on predictors of mortality, during an average period of 5 years, among 5140 adults at Tecumseh, Mich., who had been examined in 1959–60. They adopted the principles of an incidence study, and reasoned from cause to effect, but their subcohorts were not mutually exclusive, and their analyses were by non-standard methods specific to their own enquiry.

The (British) Office of Population Censuses and Surveys (OPCS) announced in 1973 plans for an exciting long-term study, known as the Longitudinal Study (LS). In due course it will have many end-points, but until recently the only publication [54] was on mortality differentials. The cohort consisted of 250,588 males and 262,484 females traced in the National Health Service Central Register; they formed 96.8% of those with four specific birth dates (1.05%) among those included in the Census taken on the night of 25/26 April 1971. As the follow-up reported was only until the end of 1975, just 4.68 years, the OPCS LS is another "atypical" cohort study. However, there had already been 13,677 deaths of males and 13,406 of females, so even the first findings [54] are not surprisingly of enormous interest. They showed, for example, that mortality was exceptionally high for children and above average for adults living in households which shared or had no access to a bath, and that mortality was above average for those living in council housing and below average for those in owner-occupied houses [35]. Those findings that bear on what is called the "healthy worker effect" (see Section 3.6) are of peculiar importance for epidemiology in occupational health. A brief summary of findings for males aged 15–64 at the time of the Census is in Table 4.

When, in 1974, McDonald *et al.* [55] published early results on mortality in a large cohort of Quebec chrysotile workers (referred to below as the "Quebec asbestos cohort"), they expressed doubts about their methods of analysis. To resolve these as far as possible, a symposium was held [56], and further exploration and appraisal of methods of analyzing a cohort led to a paper read before the Royal Statistical Society in 1977, published (with discussion) later that year as Liddell *et al.* [16]. This drew attention to several issues not always made explicit in descriptions of methods: the need to define exposure-based subcohorts, often many of them; the necessity for specifying a "study interval" over which each subject is studied in a particular facet of the analysis; the choice of reference mortality experience, particularly whether external or internal; and the importance of examining consistency of findings over different zones of the age–year space.

Liddell *et al.* [16] also recalled that, although reasoning from cause to effect in analysis of cohort studies had been predominant, argument from effect to cause (e.g. case–referent analysis within the cohort) could be highly efficient—and it must be emphasized that, at least in the

Table 4. Mortality, all causes, of males aged 15–64 by economic position, from longitudinal study of office of population censuses and surveys [54]

Economic position at Census 25/26 April 1971	Deaths between Census and 31.12.75		
	Observed	Expected	SMR
Student	26	31.5	0.83
Employed*	3021	3508.7	0.86
"Other" inactive	43	41.0	1.05
Seeking work	165	126.9	1.30
Retired (whether or not for health reasons)	91	59.4	1.53
Sick—temporarily	211	65.3	3.23
—permanently	370	94.5	3.92
All males, 15–64	3927	3927.3	1

*In a job—any work for payment or profit—in the week ended 24th April 1971, even if it was only part-time or if the person was temporarily away from work; see Ref. [54] for further detail.

present context, the two methods of reasoning have equal merit. Reviewing the use of the 1972 method of Cox [57] based on regression models (which reasons from effect to cause when a case is compared with its "risk set" [57]), Liddell *et al.* [16] proposed sampling of the risk set (incidentally matching for date of birth, and perhaps other factors). The first published results obtained from a cohort, by a case–referent form of analysis, where referents (in fact, four matched to each case) had been obtained by random sampling from the risk set, were those of G. Eyssen and the present author, quoted in Ref. [55].

Such sampling has become accepted practice. This is fortunate, because the cost of using the Cox method without sampling can be prohibitive; in the Quebec asbestos cohort, its use on five referents chosen as a random sample of the risk set for each death from selected causes cost about one-twentieth of its corresponding use on the full cohort [16]. The application by Cox [57] of his own method was for a clinical trial, and illustrates that some methods of analyzing a trial have been adapted for analyzing cohorts and vice-versa. As an early, if rather over-simplified example, see the evaluation of treatment in terms of survival described in 1937 by Hill [58].

Studies specifically designed for case–referent-within-a-cohort investigation remain rare. Perhaps the first was that described by Berry and Newhouse [59]: their cohort was of 13,460 workers at a factory producing friction materials. The only type of asbestos used was chrysotile, except during two well-defined periods before 1945 when crocidolite was also used. Their first series of cases consisted of all 10 subjects whose death could be attributed to pleural mesothelioma; four referents were matched to each individual case on five factors —sex, year of birth, survival to at least the age of death from mesothelioma, year started work in factory, and employed at factory during crocidolite period(s) for the same time as the case—in such a way that a positive association of mesothelioma with exposure to crocidolite (but not to chrysotile) could be inferred with considerable confidence. The investigations were extended to embrace lung cancer and gastro-intestinal (GI) cancer, and it was argued convincingly [59] that "the experience at this factory over a 40-year period showed that chrysotile asbestos was processed with no detectable excess mortality" [even from mesothelioma, lung cancer or GI cancer]. This important finding required detailed work histories for just 202 cases and 357 selected referents, or a total of only 559 subjects (4.2% of the cohort).

3. DATA REDUCTION FOR REASONING FROM CAUSE TO EFFECT

3.1. Life table methods

The first available description of reducing the data from a cohort study to a form amenable to comparison of subcohorts is by Frost [8]. He built up a "summation of life-experience and mortality" for each year of age (x to $x+1$), by counting the number present at the beginning of the year (l_x), and—during the year—the number added (n_x), the number withdrawn (living) (w_x), the number dying (d_x), the mean number present (L_x) and the death rate per 1000 (m_x). The derivatives from the primary figures (correcting a typographical error in the printed formulation) were:

$$l_{x+1}=l_x+n_x-w_x-d_x; \quad L_x=l_x+(n_x-w_x-d_x)/2;$$
$$m_x=(1000)d_x/L_x.$$

To study the question of morbidity and mortality in familial contacts, the 299 known contacts among the 794 present and former members of 132 negro families were identified. They were treated by making up a life-table in which the contact was entered only after the known household experience began; a second life-table included the 495 others, dating from entry into the household, *together with* any pre-exposure experience of the 299 contacts. (This reclassification of subcohorts—dealing with the problem of "transient states"—is referred to again in Section 6.) Tuberculosis attacks and deaths were counted by age, and attack rates and mortality rates were calculated by dividing by the L_x (treated as "person-years of life-experience"). Finally, the mortality and morbidity rates in the two subcohorts were adjusted to the age-distribution for the complete cohort.

3.2. The Lexis age–period space

Later authors have distinguished not only age but also period, recognizing the age–period space described by Lexis [60] (although known to actuaries at least 30 years previously [35]); see Appendix A. For each combination of sex and race, the Lexis space is subdivided into cells, such as squares of 5 years in age and/or period. This concept of subdividing the Lexis space was clearly recognized, in tabular rather than diagrammatic form, by Doll [37, 38], Case and Lea [20], and others; a diagrammatic example is in Berry [52]. Each subject can be considered as descending through the space either diagonally [6, 50, 52] or down a "staircase" [20].

3.3. Subject–years

The "subject–years method", as described by Enterline [50] and Berry [52], first follows survivors diagonally down through the Lexis space, and secondly makes use of age- and period-specific mortality rates in a reference population. Only the first stage is discussed in this section.

A subject is said to be "in view" from the time when he first comes under observation until *either*: the end of follow-up; *or* his earlier death; *or* his loss to the study (through tracing failure, migration, etc.)—see also Appendix A. [Instead of "years in view", the phrase "years of exposure" is more common, however, this can be misleading in that the "exposure" is to the possibility of death and not an indication of environmental exposure.] That subject–years in view are included up to the time any subject is lost to view, thus incorporating appropriately all available knowledge concerning such persons, is an advantage; this is not to justify high loss rates, but only to point out that the usually inevitable occasional loss can be dealt with satisfactorily.

For each subcohort, subject–years in view are accumulated for each square of the Lexis space, so that each person normally contributes to two squares in any 5-year age group and to two squares in any 5-year period; see Fig. A.1. Deaths from all causes and from specific causes are also counted square by square. (In the absence of modern computers, Doll [37, 38] had used a close approximation to the accumulation of subject–years by counting, for each single year, the numbers of men in view in each 5-year age group, adjusting for deaths within the year. Doll and Hill [6] improved even on this by taking as subject–years in any year, the average of those in view at the start and at the end of the year.)

3.4. Reference mortality experience

The term subject–years method is usually taken to refer to both stages identified in Section 3.3. Cause-specific death rates in some reference population are obtained for each square, and multiplied by the subject–years in view within the square, to yield the numbers of deaths "expected". For each cause, deaths observed and expected are summed over all squares, and the ratio of observed to expected is a Standardized Mortality Ratio (SMR), "indirect" standardization having been carried out for both age and calendar period [61]. However, it is still essential to examine the consistency or variation of SMRs in different zones of the Lexis space; see Table 4 of Ref. [50] and Table 2 of Ref. [16].

However, for comparisons of subcohorts, reference mortality is not *required*—see Appendix B. For instance, Doll and Hill [6] compared mortality experience between subcohorts defined in terms of amount of tobacco smoked, within their complete cohort; they did use reference information, but solely an age-distribution, and only for further reduction of data. The essential remains the determination, square by square in the Lexis space, of numbers of deaths and of subject-years, for each subcohort. These could be used to obtain age-, period- and exposure-specific rates (not ratios) of mortality, as exploited by Frome [62]; see also Appendix B. Even further freedom could be introduced by

use of a rate for every 1-year square in the Lexis space. With some diseases—tuberculosis, lung cancer, stomach cancer, for instance—secular changes in mortality rates have been so rapid that this approach may occasionally be justified.

3.5. *The variant of Case and Lea*

Although it is often stated that Case and Lea [20] introduced the subject–years method, not only had Doll [37] employed its fundamentals earlier but—more important—the Case/Lea method is a decided variant. After the Lexis space had been subdivided into 5-year squares, subjects were followed down the "staircase". For example, at the start of 1931, there were 387 survivors of cohort (i) aged 35–39; in the next 5 years, 22 died leaving 365 survivors, now aged 40–44 at the start of 1936; by the end of 1950, there had been 111 deaths, leaving 260 survivors into 1951. For each of the seven initial age groups (25–59), there were six steps down—but with imperfections due to the short marginal periods (1930 and 1951–52). The next stage was to note the number of deaths, in each 5 years, from all causes [and from certain cancers] occurring to men in each age group at the start of each date period. Then the number of men in each such group was multiplied by the appropriate ${}_5q_x$ value from the relevant Life Table for England and Wales (adjusted for the marginal periods), to provide the "expected" number of deaths in each square. Thereafter, the number of deaths, found and expected, were summed over dates and then ages, to yield the totals of Table 3. (It should be noted that tests had demonstrated marked inconsistency among age groups of the ratios of deaths found to those expected.)

It can be shown that this variant is only an approximation to the true—and optimal (see Section 4)—subject–years method. It is perhaps fortunate that the variant has not been generally followed, but in its day it had the great advantage that it did not require electronic computing.

3.6. *Choice of reference mortality experience*

Where the analysis of cohort studies does demand the mortality experience of a reference population, its selection remains a matter of considerable concern [63], and there is seldom any means of checking its suitability. Even the check made by Case and Lea [20], by means of their cohort (iii), was "necessary but not sufficient". If amputees had had excess cancer mortality, considerable doubt must have been felt about all the findings; even when they did not, there was no guarantee that the reference was truly appropriate for cohorts (i) and (ii).

According to Fox and Collier [64], William Ogle in his *Letter to the Registrar-General on mortality in the registration districts of England during the* 10 *years* 1851–60 had pointed out that there must exist what is now called the "healthy worker effect"—men who can work are in general healthier than those who cannot—and that this would differ from industry to industry, because of the varying physical demands of different occupations. In 1975, Goldsmith [65] had written that several recent studies had sought to compare observed mortality in some employed group with that in the general population; he emphasized the bias inherent in comparing employed populations with general population mortality, because of the "healthy worker effect", and suggested that occupational cancer risks thereby tended to be underestimated. There were three replies, but no truly realistic solutions.

Fox and Collier [64] named three factors as of importance: the selection of healthy persons for employment, the survival in the industry of healthier persons, and the length of time the working population had been followed. At first reading, the findings of these authors appear to demonstrate major selection and survival effects. However, interpretation is not always clear, because of inadequate definitions, particularly of age groups. The estimates of the "Healthy Worker Effect" were so extreme (SMR, all causes, all ages, of 0.37 for men within 5 years of first employment), and the findings in relation to survival in industry so unsatisfactory (because past employees were not classified as to whether they were still at work or not), that it is almost impossible to accept these "results" as evidence of selection in the usual sense of this word. Medical colleagues have pointed out that if any method of selection—medical, social, personal, or industrial—could achieve such astounding improvements in mortality as those reported by Fox and Collier [64], not only among the comparatively elderly, and be maintained (at lower but still substantial levels) for many years after first employment, it was surprising that no member of their profession knew of the method.

Doll [44] has expressed the view that (rather less extreme) selection effects may persist for a long time for mortality from many causes, but that they cannot be expected to survive for as

long as 5 years in relation to cancer. Meanwhile, Table 4 shows results that are in full accord with common sense.

External and internal reference mortality experience. The Quebec asbestos cohort was sufficiently large, comprising nearly 11,000 men, and with a wide enough range of measures of exposure that exposure–response could be examined in relation both to an "external" reference, i.e. the mortality experience of the population of the Province of Quebec, and to the "internal" reference that consisted of the mortality experience of the entire cohort. Liddell *et al.* [16] showed that the *relative* gradients of the curves were virtually identical.

Advantages of the experience of the entire cohort as internal reference are several. First, and of great importance, certain factors—geographical location, socio-economic class, social environment and habits, etc.—all pertinent [63], are minimized or even overcome. Secondly, relationships can be examined even for exposure-related diseases that are rare enough that expectations derived from mortality in the general population are too low to be stable. Thirdly, the otherwise dubious habit of re-coding cause of death to obtain more accuracy can also be permitted; this would not be so if the reference were external. (An exception arose in 1955: Doll [38] fully accepted that it is a dubious procedure to include autopsy cases in numerators but not in denominators, but justified the procedure in his study by an argument of some subtlety, which concluded that for the excess number of lung cancer cases to be reasonably attributed to chance, it would be necessary for the expected cases to be nearly 6 times the number estimated on the highest population rates—and that would have meant that lung cancer would have had to be capable of detection in over one-fifth of all men at death throughout 1931–1951.) This third advantage is important when pathological evidence can be acquired [38, 40, 41, 43, 66, and others].

4. METHODS OF ANALYSIS BY CAUSE TO EFFECT REASONING

Berry [52] has shown that the assumptions which have been implicit over the years in the use of the subject–years method, when making use of reference mortality, are all justified asymptotically; in particular it is valid to treat observed numbers of deaths as Poisson variables, while the comparison of observed and expected for all (or particular) causes of death gives a valid measure of excess risk. (Hartz *et al.* [67] claimed, in a paper accepted before [52] was published, that these procedures were not sound. Later discussion, in four dissenting "Letters to the Editor" and two "Authors' Replies", culminated in a review [68] which stressed that "the subject–years method is optimal for the purpose for which it is primarily intended: statistical inference for relative death rates".)

Breslow's 1985 chapter [17] reviewing the analysis of (reduced) data from cohort studies—effectively comparing the mortality experience of subcohorts, defined in terms of exposure—is up-to-date and comprehensive. (It has 73 references, 44 published after the 1977 appraisal of methods [16], 29 of these in the years 1981–85; only a handful of the 29 earlier references were specifically on methods.) It would thus be gratuitous to attempt another review here. Instead, the following paragraph is an extension of Breslow's summary [17], with some changes of wording mainly to accord with usage above.

An efficient method of analysis, reasoning from cause to effect, is to fit Poisson regression models to grouped data, when the data comprise disease cases and subject–years of observation, both classified not only by discrete categories of age and period, but also by various aspects of exposure. The models can be extended for use when disease rates and also exposure variables vary continuously with age or time; such extension leads to the well-known proportional hazards model [57]. Further, this extension for use with age- and time-variable exposures does not require any grouping of the exposure variables. In *multiplicative* (or relative risk) situations, incorporation of external reference mortality experience is unlikely to improve greatly the estimates of exposure effects. Nevertheless, because reference mortality (or morbidity) experience may not apply strictly to the cohort under study, an (unknown) scale factor can be introduced to adjust the reference rates so that they more nearly represent the true background rates. In *additive* (or excess risk) models, however, external reference rates are more likely to improve estimates.

Also discussed [17] are choices between models, and problems specific to fitting non-multiplicative relationships. Some essential differences between models are explained in Appendix B, below. Much more detail is given in **The Design and Analysis of Cohort Studies** [69].

5. ANALYSIS BY REASONING FROM EFFECT TO CAUSE

In most "typical" cohorts, i.e. of large population groups over many years, the use of methods of analysis stemming from that of Cox's 1972 paper [57] would be inordinately expensive, unless there is sampling of the risk set for each case [16, 17]. Random sampling from each risk set, followed by case–referent analysis, has been shown empirically [16] and theoretically [16, 70] to provide unbiased estimates of the hazards ratio. The criteria for selection of referents will, inevitably, involve important decisions: Lubin and Gail [70] have shown that what might seem to be sensible restrictions on the choice of referents can lead to appreciable bias; care must be taken to allow a referent, once selected, to be included as such even if he later becomes a case or has previously been selected as referent for an earlier case [71]. However, once cases have been identified and referents chosen, the analysis can follow standard lines, although the particular form of analysis will depend on the way in which referents have been sampled. As methods of case–referent analysis were fully reviewed in 1980 by Breslow and Day [22] and in 1982 by Breslow [72], they are not within the purview of the present paper.

6. STUDY INTERVALS

A "study interval" has already been introduced (Section 2.4, Ref. [16]) as that specified period over which each individual subject is studied in a particular facet of the analysis. Selection of an appropriate study interval for analysis that reasons from cause to effect (i.e. for comparing mortality in subcohorts defined by their exposures) has been discussed by, among others, Enterline and Henderson [51] and Fox [73], who pointed out that long exposure appears to *improve* life expectancy. Liddell *et al.* [16] proposed one way of comparing like with like in occupational cohorts: to make the study interval for analysis start for all men at the same point relative to entry to employment; for example, the subcohort with at least 20 years of service was contrasted with subcohorts who had also survived at least 20 years since their service began, although the service itself was shorter. (A lag after first exposure is frequently desirable because of the often long interval before the manifestation of disease resulting from the exposure. Thus, where the interval between first exposure and manifestation of disease is thought to be of the order of 30 years, disease arising within say a dozen years of first exposure is most unlikely to be attributable to the exposure. [A well-known example is the occurrence of pleural mesothelioma after exposure to amphibole asbestos, where the shortest interval observed has been at least 18 years, with an average closer to 35 years.] Earlier deaths can be examined separately; but then there is the proviso that the study interval must terminate at the end of the lag—in order that the examinations can be independent, i.e. orthogonal.)

In this approach, it follows that, to investigate the effects of exposure variables which vary with time, it is essential to base the classification of subcohorts on the variables at the start of the study interval. Otherwise, a subject's classification will change from time to time *during* the study interval. This problem of "transient states" was overcome by Frost [8] (and others before him)—see Section 3.1—and by Mantel and Byar [74], who re-introduced the idea of reclassifying each subject as he crossed a boundary between levels of classification. Then each individual may contribute years in view to more than one of the subcohorts during the course of the study.

7. MEASURES OF EXPOSURE

What "index" of exposure to adopt is often a moot point. It is important to make the most of whatever measurements, usually from the past, are available. Factors to be considered include the age of the subject when his exposure commenced (or occurred), its duration (with or without gaps), and its intensity (which may well have fluctuated greatly). For one-shot exposures, the choice of index is simplified, as duration and fluctuation are usually irrelevant or cannot be estimated; nevertheless, intensity can be of vital importance, as in the "natural experiment" of the atomic bomb explosions, where distance from the epicentre has been used as an inverse indication of intensity of radiation dose. For many agents, accumulated exposure (the integral of duration multiplied by intensity), such as "pack-years of cigarettes smoked", has been adopted. Because many diseases of particular interest, such as lung cancer, are probably not influenced by any exposure in the last few years of life, an index has often been adopted with exposure accumulated only up to

some (arbitrarily chosen) cut-off point several years before death. Other possibilities exist; for example, exposure can be weighted according to when it was suffered, the higher weights usually being given to exposures early in life. Cut-off or weighting can be introduced comparatively easily into case–referent analysis, but are less easy to impose in analysis reasoning from cause to effect.

Occasionally, the available information on exposure can be manipulated in such a way that different indices suggested by various concepts of disease processes can be examined. This was possible in the Quebec asbestos cohort [75]: for each of the 11,000 subjects, estimates had been made of what proportion of each year (1904 through 1966) had been worked in the asbestos industry and of the average intensity of exposure in the year. It had been shown [16, 75] that the relation between accumulated exposure to asbestos and the relative risk of lung cancer was effectively linear (without threshold), but this did not appear particularly plausible in biological terms—although similar patterns had been identified in other studies [76]. A painstaking search [77] failed to find an index of exposure that correlated better. Meanwhile, it had been demonstrated beyond reasonable doubt that a measure of exposure based on duration alone—a possible index—was less well correlated with relative risk of lung cancer than when intensity of exposure was incorporated into the index [75 Table 7(d), 76].

Measurements of exposure are subject to error, often large. Even if such error is itself unbiased, it may give rise to bias (usually towards zero) in an estimate of slope of an exposure–response relationship. Armstrong and Oakes [78] indicated how this bias could be estimated, on certain more-or-less realistic assumptions, in case–referent analysis of a cohort mortality study; for data from the Quebec asbestos cohort [75], they suggested that the original estimate of slope [79] was too shallow by about 15%, a deviation of roughly 0.5 SE, yet "small in comparison with the other uncertainties involved in this type of study". Doll and Peto [80] stated that the general effect of random errors in dosimetry was likely to make the relationship between risk and measured dose less extreme, in other words to bias a higher-powered relationship with dose (e.g. incidence proportional to the square of the "true insult") into a lower-powered one (e.g. incidence directly proportional to recorded exposure). D. C. Thomas (personal communication) reported a "sub-linear" relation of lung cancer risk to amount of cigarettes smoked in the study described by Liddell *et al.* [81], and speculated that this was a reflection of a linear relationship biased to lower power by errors in the quantification of amount smoked.

When exposures are evaluated serially, from annual medical examinations for example, they become repeated multi-variate measurements and different concerns arise. Tango [82] fitted models of the *change* in measurements prior to the onset of disease; for this to be possible, however, strong assumptions had to be made about the joint distributions of the measurements.

8. SOME FUNDAMENTALS IN THE DESIGN OF A COHORT STUDY

8.1. Composition of the cohort

The simplest, but very important, cohort consists of all who have been subjected to certain exposures; examples have already been cited [26–28, 39–41]. Clearly, such cohorts usually comprise persons with a wide range of dates of birth. So also did the cohort surveyed by Frost [8] in 1933.

There are however good reasons why cohort surveys should be based on goups defined by dates of birth, rather than on exposure factors, but there are few such surveys; one is the Quebec asbestos cohort [75]. This consisted of all 11,379 persons born 1891 through 1920 who worked at least a month in the asbestos production industry. Thus, it was comparatively homogeneous in selection and retention factors, avoiding any effects of premature selection *out* of the industry, on health grounds or otherwise, of those who stayed in it for more than a month; however, it still could not overcome "natural" selection *into* different exposures.

Such follow-up has been rare; most occupational cohorts are based on a cross-section of those at work at a particular time, or on a "slice" (thinner or thicker according to definition) such as those employed since a specific date, or for at least a given length of time (with corresponding definitions for environmental measures). Such "quasi-cross-sectional starts" must surely lead to some ungeneralizability of findings, because of selection factors related to period of employment or other exposure [83]. In the one report known to the author on how such choices, rather than of a birth cohort, affect the

main results, Oakes and McDonald [84] provided some empirical support for two of three common restricted choices of occupational cohort when subjected to age-matched case–referent analysis; they suggested that a restricted choice may be useful—but only when complete follow-up of an entire birth cohort is not feasible.

A decision on cohort definition may also depend on the investigator's expectations of the ease or difficulty of tracing (which usually depend on age, calendar period and other factors) and on beliefs in the length of latency, and in the importance of other suspected factors in the natural history of the disease(s) of particular interest. The possibility must be examined that a decision made in the light of considerations such as these may lead to biases due to selection factors.

It is important that a cohort need not be a large population group. Indeed, the relationships between rare diseases and certain exposures with high RRs can sometimes best be examined by studies of necessarily small cohorts. The investigations of mesothelioma in former gas mask assemblers [40, 41] are cases in point, and Appendix C illustrates how reasonable the approach was. Indeed, it is difficult to see how this phenomenon could have been investigated except with a cohort design. There have also been numerous studies of small cohorts, usually because the exposed population was strictly circumscribed. If such a study has little power, it may nevertheless suggest the need for finding another larger group for further investigation.

8.2. Follow-up period

As has been seen, the follow-up period need not always be of many years; a prime example is the series of studies of rubella in pregnancy [21], where only a cohort approach was appropriate for assessing the risk of congenital defect, as explained by Doll [1]. It is, of course, essential that the follow-up be long enough to permit the development of detectable disease outcomes that may be the results of exposure to the factor(s) under investigation.

8.3. Sampling of referents

Although the sampling of referents for the determination of their exposures has been practised since 1955 [43], it was only after about 1977 that the analysis of a cohort by case–referent methods became common. Even now, there have been few attempts to incorporate this analytical approach into the design of the study [59]. The potential advantage is that, although the complete cohort has to be followed in traditional fashion (so that the recently coined term "nested case–control study" seems to have no superiority), one need determine the exposures only of the cases and of the selected referents. The potential will not be achieved unless it is possible to avoid bias in evaluating exposures [85]; that this is at least occasionally possible has been shown by J. C. McDonald (who planned in 1958 that sera from all subjects [pregnant women] would be frozen, but that only those for cases [women with abnormal outcome of pregnancy] and a random sample of referents—a total of only 20% of the complete cohort—would be analyzed) reported in [86], and Armitage [85] who emphasized success in the Quebec asbestos cohort [16]. Some subtle difficulties were mentioned by Liddell [87]: reliable evidence had been accumulated of a tendency to understate the exposures of referents (but not of cases). To "correct" for this recording bias would have reduced the slopes of the exposure–response lines by about 15%—but they were already believed [78] to have been reduced to a similar extent because of random error. Thus to remove the bias from the recorded exposures might have led to an estimate of the exposure–response relationship even more seriously biased (towards the null). The issue was raised in public, but in a lively debate there was no agreement on the solution.

In one recent design [88], a subset of the cohort is *randomly* selected for ascertainment of exposures, while exposure histories are assembled for other cohort members only if disease develops; this, of course, means that—as above—the complete cohort has to be followed to ascertain cases. Possible advantages over sampling from each case's risk set have not yet been fully evaluated, but the design allows for internal comparisons according to Prentice [88], and also, as Boivin and Wacholder [89] show, for external comparisons. The best name for this design seems to be "case–minicohort"; the term "case–cohort" [88] is rather less satisfactory here, because it describes better the designs of Farr and Snow [27, 28].

It remains esential to complete the follow-up of the cohort and to avoid bias in the evaluation of exposures for both cases and sampled referents. The latter means that the evaluation must be done truly "blind", and hence at one time—

which cannot be before all the cases have been identified. However, exactly how the analysis is carried out will depend on the way in which the referents have been sampled: they may have been matched to the cases for date of birth (and perhaps other factors); if not, the samples may have come from some stratified or non-stratified random procedure; etc.

8.4. Other variations in design

The incorporation into the study of a second cohort as reference has been attempted. Unless the reference cohort has been matched to the study cohort [90], it must be of doubtful relevance. The use of any second cohort renders the "SMRs" much less stable, and can introduce other major problems.

In another interesting design, pairs of exposed and non-exposed subjects, matched for other factors, are followed at the same time [91].

9. FUTURE DEVELOPMENTS

Suggestions by Breslow [17] for future methodological research included: further development of the excess risk (additive) model for continuous data; and consideration of the contribution of external reference rates for models other than the multiplicative. The present author feels it would be particularly interesting to investigate the utility of reference rates from an internal source, probably the entire cohort. Breslow [17] also suggested re-evaluation of the widely held notion that only four or five referents per case are needed for case–referent analyses. They do suffice for most cohort studies; however, more referents may be of value for estimating *high* relative risks (above about 3) associated with *rare* exposures among the referents (less than 20% exposed) [92]. Breslow [17] proposed study of variation of the referent/case ratio according to the case's exposure.

After the analysis, what? Interpretation of published material is difficult if reliance has been placed on apparently inappropriate external reference rates. This is of particular importance when trying to compare exposure–response relationships in studies of different cohorts where the responses have been expressed only as observed and expected cases (or as observed cases and SMRs) in relation to an external reference. A means of tackling the problem where the SMRs for the minimally exposed are not close to unity has been proposed [76, 93], and adopted for lung cancer in relation to asbestos exposure by several workers, including Acheson and Gardner [94]. (This interpretative approach might well have been unnecessary had the analyses either incorporated "scale factors" or been in relation to an internal reference—see Sections 4 and 3.6.)

The question above also demands thought on how results of analysis should be presented. Berry [95] calculated reductions in the expectation of life, which gave a more comprehensible summary of the implications of elevated death rates than did SMRs. A possible improvement in presentation might be to calculate the excess of Potential Years of Life Lost (PYLLEX), based on the appropriate reference life expectancy, and exploit the desirable PYLLEX property of additivity over causes [96]; investigation is needed before adoption for cohort studies, although approximate standard errors are now available. One possible weakness of PYLLEX is that it may be positive when the corresponding SMR is less than unity, and vice-versa.

10. CONCLUDING REMARKS

Cohort studies are considerably more flexible than has often been realized. The mortality experience of an external reference population is of course required when the cohort cannot be subdivided into several subcohorts with varying degrees of exposure to the factor(s) under investigation. Otherwise—when the main objective is to compare the subcohorts among themselves, and not each, separately, against an external reference—such reference serves little purpose, in the reviewer's opinion, other than to place the mortality of the entire cohort in some general context. If, however, such an attempt to place the mortality of the cohort—or of several subcohorts—in the general context of the population is deemed essential, the use of the subject–years method has recently been shown to be entirely appropriate.

The SMR is undoubtedly the most commonly used analytic tool, but there is heated debate as to its general utility—because of inevitable doubts about the validity of the reference mortality rates—and as to the possibilities of comparing more than one SMR in the same study (say to investigate exposure–response relations), as well as to other possibilities of misuse [97]. Nevertheless, there may be no practical alternative.

Certain problems over study intervals, exposure measures, etc. are now fully recognized

and are often capable of satisfactory resolution. In the subject–years method, it has become possible to incorporate "scale factors" for adjusting reference mortality rates so that they more nearly represent true background rates. If scale factors have not been used, recent methods of interpretation allow for the fact that the members of a subcohort who have been exposed minimally (or not at all) to the factor(s) under consideration frequently yield SMRs that differ considerably from unity. These methods permit some comparison of the mortality of exposed subcohorts against that of the unexposed.

A comparatively modern gain of major importance lies in the realization that a considerable proportion of the great expense of the study of a large cohort for a long period can be eliminated by the evaluation of exposures for only a sample of referents for any particular series of cases, with only slight loss of statistical power. Even if the cost of collecting unnecessary environmental information cannot be avoided—because of potential bias such as could be introduced by evaluating exposures of the referents at the start of the study, and of the cases at the end—at least extremely heavy computing costs can be reduced by an order of magnitude. Nevertheless, it is important that in these designs (or, less comprehensively, in these methods of analysis) possible biases are eliminated. Meanwhile, further work is required on the best method(s) of sampling the referents.

Further flexibility lies in the fact that the cohort design is not confined to following *large* population groups over *long* periods. It used to be—and still is—believed that the case–referent design is the "natural" choice for rare diseases (although the rarity of the disease has in recent years ceased to be a *sine-qua-non* for case–referent studies). However, the relationships between rare diseases and certain exposures with high RRs can sometimes *best* be examined by cohort studies, as explained in Section 8.1. Such studies may well be quite small; and comparatively short follow-up has been exemplified in [21, 26–28, 30, 45 and 54].

In conclusion, the cohort study in epidemiology is not only alive but flourishing. Its success owes much to the inventiveness of the researchers mentioned in this review, and of course to many others, including the biostatisticians who have developed the ideas of Sir David Cox, and their improvements, into a coherent "schema". No process will ever attain perfection, and improvements will still arise, but we already have, in the cohort study, an extremely powerful set of tools at the service of epidemiology.

Acknowledgements—I am particularly grateful to the International Statistical Institute for permission to make use of material which appeared in the ISI Bulletin. This review is based on an invited paper [98] for the Centenary Session of the ISI, held in Amsterdam during August 1985, but much has been added, and the text has been largely rewritten. I also thank many colleagues, epidemiologists and biostatisticians, in several countries, whose encouragement and advice have been invaluable. In particular, Corbett McDonald—whose name appears so frequently in the bibliography as to show the importance of his many contributions in the field reviewed—helped enormously in the early stages, when I was floundering towards a coherent presentation. Norman Breslow invited me to present the "overview" at the Amsterdam Session of the ISI, and was particularly helpful in providing advance copies of important material, including Ref. [17]; this made my task quite surprisingly easier than otherwise, as I could omit detailed description of analytical methods and refer to Breslow's own inimitable offerings. Sholom Wacholder has been an enthusiastic supporter of the project since he first saw a draft in the Spring of 1985; his contributions have been of great benefit. Bernard Benjamin has also provided much positive input [35], from his own especially informed viewpoint. Finally, Richard Doll, whose insights and drive have led to so many advances in the general field, has always been an inspiration, and he has provided many personal communications, which I have incorporated as Ref. [44], and for which I am especially grateful. His letter of review of an earlier draft—courteous, critical and highly constructive—provided great encouragement when it was needed most.

REFERENCES

1. Doll R. Retrospective and prospective studies. In: Witts LJ, Ed. **Medical Surveys and Clinical Trials: Some Methods and Applications of Group Research in Medicine.** London: Oxford University Press; 1964: 2nd edn, pp. 71–98.
2. Last JM, Ed. **A Dictionary of Epidemiology.** New York: Oxford University Press; 1983.
3. Cochrane AL. **Effectiveness and Efficiency.** Abingdon: Nuffield Provincial Hospitals Trust; 1972.
4. Truelove SC. Follow-up studies. In: Witts LJ, Ed. **Medical Surveys and Clinical Trials: Some Methods and Applications of Group Research in Medicine.** London: Oxford University Press; 1964: 2nd edn, pp. 99–114.
5. Hill AB. **A Short Textbook of Medical Statistics.** London: Hodder & Stoughton; 1984: 11th edn, 272–277.
6. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking: a second report on the mortality of British doctors. **Br Med J** 1956; 2: 1071–1081.
7. White C, Bailar JC III. Retrospective and prospective methods of studying association in medicine. **Am J Publ Health** 1956; 46: 35–44.
8. Frost WH. Risk of persons in familial contact with pulmonary tuberculosis. **Am J Publ Health Nation's Health** 1933; 23: 426–432.
9. Springett VH. What is a cohort? **Br Med J** 1979; 1: 126.
10. Jacobs AL. What is a cohort? **Br Med J** 1979; 1: 266.
11. Frost WH. Letter to the late Dr. Edgar Sydenstricker, 1935. **Am J Hyg** 1939; 30: 91–92.

12. Frost WH. The age selection of mortality from tuberculosis in successive decades. **Am J Hyg** 1939; 30: 91–96.
13. Korteweg R. The age curve in lung cancer. **Br J Cancer** 1951; 5: 21–27.
14. McDonald JC, Springett VH. The decline of tuberculosis mortality in Western Europe. **Br Med Bull** 1954; 10: 77–81.
15. Alderson M. **An Introduction to Epidemiology.** London: Macmillan; 1976: 27–29.
16. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: appraisal by application to asbestos mining. (With discussion.) **J R Stat Soc A** 1977; 140: 469–491.
17. Breslow NE. Cohort Analysis in Epidemiology. In: Atkinson AC, Fienberg SE, Eds. **A Celebration of Statistics: The ISI Centenary Volume.** New York: Springer-Verlag; 1985: 109–143.
18. Kupper LL, Janis JM, Karmous A, Greenberg BG. Statistical age–period–cohort analysis: a review and critique. **J Chron Dis** 1985; 38: 811–830.
19. Holford TR. An alternative approach to statistical age–period–cohort analysis. **J Chron Dis** 1985; 38: 831–836.
20. Case RAM, Lea JA. Mustard gas poisoning, chronic bronchitis, and lung cancer. **Br J Prev Soc Med** 1955; 9: 62–72.
21. Hill AB, Doll R, Galloway TM, Hughes JPW. Virus diseases in pregnancy and congenital defects. **Br J Prev Soc Med** 1958; 12: 1–7.
22. Breslow NE, Day NE. **Statistical Methods in Cancer Research I: The Analysis of Case–Control Studies.** Lyon: International Agency for Research on Cancer; 1980.
23. Susser M, Adelstein A. Introduction to Humphreys NA, Ed. **Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr.** London: Sanitary Institute; 1885. (Reprinted Metuchen NJ: Scarecrow; 1975.)
24. Louis PCA. **Recherches sur les Effets de la Saignée dans quelques Maladies inflammatoire.** Paris: Baillière; 1835.
25. Semmelweis IP. **The Etiology, the Concept and the Prophylaxis of Childbed Fever.** Pest, Vienna and Leipzig: Harleben's Verlags-Expedition; 1861. (Translated by RP Murphy, and reprinted in **Medical Classics** 1941; 5: 350–772.)
26. Farr W. **The Mortality of Cholera in England, 1848–49.** In: Humphreys NA, Ed. **Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr.** London: Sanitary Institute; 1885: 333–351.
27. Farr W. 17th Annual Report. In: Humphreys NA, Ed. **Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr.** London: Sanitary Institute; 1885: 357–363.
28. Snow J. **On the Mode of Communication of Cholera.** London: Curchill; 1854: 2nd edn.
29. Terris M. The changing relationships of epidemiology and society: The Robert Cruikshank Lecture. **J Publ Health Policy** 1985; 5: 15–36.
30. Terris M, Ed. **Goldberger on Pellagra.** Baton Rouge: Louisiana State University; 1964.
31. Population Investigation Committee. **Maternity in Great Britain.** London: Oxford University Press; 1948.
32. Atkins E, Cherry N, Douglas JWB, Kiernan KE, Wadsworth MEJ. The 1946 British cohort: an account of the origins, progress and results of the National Survey of Health and Development. In: Mednick SA, Baert AE, Eds. **Prospective Longitudinal Research: an Empirical Basis for the Primary Prevention of Psychosocial Disorders.** Oxford: Oxford University Press; 1981: 25–30.
33. Sörbom D. A statistical model for analysis of longitudinal data: LISREL. In: Mednick SA, Baert AE, Eds. **Prospective Longitudinal Research: an Empirical Basis for the Primary Prevention of Psychosocial Disorders.** Oxford: Oxford University Press; 1981: 16–19.
34. Mednick SA, Baert AE, Eds. **Prospective Longitudinal Research: an Empirical Basis for the Primary Prevention of Psychosocial Disorders.** Oxford: Oxford University Press; 1981.
35. Benjamin B. Personal communications from Professor Bernard Benjamin to the author; 1987.
36. Shepherd PM. **The National Child Development Study.** London: The City University Social Statistics Research Unit; 1985: NCDS Working Paper No. 1.
37. Doll R. The causes of death among gas-workers with special reference to cancer of the lung. **Br J Ind Med** 1952; 9: 180–185.
38. Doll R. Mortality from lung cancer in asbestos workers. **Br J Ind Med** 1955; 12: 81–86.
39. Beebe GW. The atomic bomb survivors and the problem of low-dose radiation effects. **Am J Epidemiol** 1981; 114: 761–783.
40. Jones JSP, Smith PG, Pooley FD, Berry G, Sawle GW, Madeley RJ, Wignall BK, Aggarwal A. The consequences of exposure to asbestos dust in a wartime gas-mask factory. In: Wagner JC, Ed. **Biological Effects of Mineral Fibres,** Vol.2. Lyon: International Agency for Research on Cancer; 1980: Vol. 2, 637–653.
41. McDonald AD, McDonald JC. Mesothelioma after crocidolite exposure during gas mask manufacture. **Environ Res** 1978; 17: 340–346.
42. Doll R. Lung cancer and cigarette smoking. **Acta UICC** 1959; 15: 417–423.
43. Court Brown WM, Doll R. **Leukaemia and Aplastic Anaemia in Patients Irradiated for Ankylosing Spondylitis.** London: Her Majesty's Stationery Office; 1957: Medical Research Council Special Report Series 295.
44. Doll R. Personal communications from Sir Richard Doll FRS to the author; 1985–1987.
45. McDonald AD. Maternal health and congenital defect: a prospective investigation. **N Engl J Med** 1958; 258: 767–773.
46. Kagan A, Dawber TR, Kannel WB, Rovotskie N. The Framingham study: a prospective study of coronary heart disease. **Fedn Proc** 1962; 21 (Suppl. 11): 52–57.
47. Cornfield J. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. **Fedn Proc** 1962; 21 (Suppl. 11): 58–61.
48. Mantel N. Synthetic retrospective studies and related topics. **Biometrics** 1973; 29: 479–486.
49. Kupper LL, McMichael AJ, Spirtas R. A hybrid epidemiological study design useful in estimating relative risk. **J Am Stat Assoc** 1975; 70: 524–528.
50. Enterline PE. Mortality among asbestos products workers in the United States. **Ann NY Acad Sci** 1965; 132: 156–165.
51. Enterline PE, Henderson V. Type of asbestos and respiratory cancer in the asbestos industry. **Arch Environ Health** 1973; 27: 312–317.
52. Berry G. The analysis of mortality by the subject-years method. **Biometrics** 1983; 39: 173–184.
53. Higgins MW, Keller JB. Predictors of mortality in the adult population of Tecumseh. **Arch Environ Health** 1970; 21: 418–424.
54. Fox AJ, Goldblatt PO. **Longitudinal Study: Socio-demographic Mortality Differentials.** London: Her Majesty's Stationery Office; 1982: Series LS no. 1.

55. McDonald JC, Becklake MR, Gibbs GW, McDonald AD, Rossiter CE. The health of chrysotile asbestos mine and mill workers of Quebec. **Arch Environ Health** 1974; 28: 61–68.
56. Liddell FDK. Occupational mortality in relation to exposure. **Arch Environ Health** 1975; 30: 266–267.
57. Cox DR. Regression models and life tables. **J R Stat Soc B** 1972; 34: 187–202.
58. Hill AB. Principles of medical statistics. XI. Life tables and survival after treatment. **Lancet** 1937; 1: 646–648.
59. Berry G, Newhouse ML. Mortality of workers manufacturing friction materials using asbestos. **Br J Ind Med** 1983; 40: 1–7.
60. Lexis W. **Einleitung in die Theorie der Bevölkerungs-Statistik** Strasbourg: Trulner; 1875. (Cited in Wilson A, Rees R. Spatial demography: some comments. **Statistician** 1976; 25: 59–61.)
61. Liddell FDK. The measurement of occupational mortality. **Br J Ind Med** 1960; 17: 228–233.
62. Frome EL. The analysis of rates using Poisson regression models. **Biometrics** 1983; 39: 665–674.
63. Gardner MJ, Ed. **Expected Numbers in Cohort Studies. Proceedings of a Meeting held on 21st May 1984 at the MRC Environmental Epidemiology Unit.** Southampton: Medical Research Council Environmental Epidemiology Unit; 1984.
64. Fox AJ, Collier PF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. **Br J Prev Soc Med** 1976; 30: 225–230.
65. Goldsmith JR. What do we expect from an occupational cohort? **J Occup Health** 1975; 17: 126–127.
66. McDonald JC, McDonald AD, Gibbs GW, Siemiatycki J, Rossiter CE. Mortality in the asbestos mines and mills of Quebec. **Arch Environ Health** 1971; 22: 677–686.
67. Hartz AJ, Giefer EE, Hoffmann RG. A comparison of two methods for calculating expected mortality. **Stat Med** 1983; 2: 381–386.
68. Keiding N, Vaeth M. Calculating expected mortality. **Stat Med** 1986; 5: 327–334.
69. Breslow NE, Day NE. **Statistical Methods in Cancer Research II: The Design and Analysis of Cohort Studies.** Lyon: International Agency for Research on Cancer; 1987.
70. Lubin JH, Gail MH. Biased selection of controls for case–control analyses of cohort studies. **Biometrics** 1984; 40: 63–75.
71. Robins JM, Gail MH, Lubin JH. More on "biased selection of controls for case–control analyses of cohort studies". **Biometrics** 1986; 42: 293–299.
72. Breslow NE. Design and analysis of case–control studies. **Ann Rev Publ Health** 1982; 3: 29–54.
73. Fox AJ. Vinyl chloride and mortality? **Lancet** 1976; 2: 417.
74. Mantel N, Byar DP. Evaluation of response–time data involving transient states: an illustration using heart–transplant data. **J Am Stat Assoc** 1974; 69: 81–86.
75. McDonald JC, Liddell FDK, Gibbs GW, Eyssen GE, McDonald AD. Dust exposure and mortality in chrysotile mining, 1910–1975. **Br J Ind Med** 1980; 37: 11–24.
76. Liddell FDK, Hanley JA. Relations between asbestos exposure and lung cancer SMRs in occupational cohort studies. **Br J Ind Med** 1985; 42: 389–396.
77. Thomas DC. Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. **Scand J Work Environ Health** 1983; 9: 353–366.
78. Armstrong BG, Oakes D. Effects of approximation in exposure assessments on estimates of exposure–response relationships. **Scand J Work Environ Health** 1982; 8 (Suppl. 1): 20–23.
79. Berry G. Dose response in case–control studies. **J Epidemiol Commun Health** 1980; 34: 217–222.
80. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationship among regular smokers and lifelong non-smokers. **J Epidemiol Commun Health** 1978; 32: 303–313.
81. Liddell FDK, Thomas DC, Gibbs GW, McDonald JC. Fibre exposure and mortality from pneumoconiosis, respiratory and abdominal malignancies in chrysotile production in Quebec, 1926–75. **Ann Acad Med Singapore** 1984; 13 (Suppl.): 340–344.
82. Tango T. Statistical model of changes in repeated multivariate measurements associated with the development of disease. **Computational Stat Data Anal** 1985; 3: 77–88.
83. Liddell FDK, McDonald JC. Survey design and analysis. In: McDonald JC, Ed. **Recent Advances in Occupational Health.** Edinburgh: Churchill Livingstone; 1981: 95–105.
84. Oakes D, McDonald JC. Restricted cohort study designs. **Scand J Work Environ Health** 1982; 8 (Suppl. 1): 30–33.
85. Armitage P. In: Discussion of the Paper by Professors Liddell and McDonald and Dr Thomas. **J R Stat Soc A** 1977; 140: 488.
86. Stewart JA. **A Multivariate Statistical Approach to the Problem of Infections During the Early Months of Pregnancy and Their Relationship to Abortion, Still Birth, Congenital Malformations, and Neonatal and Infant Mortality.** Montreal: McGill University; 1974: MSc Thesis.
87. Liddell FDK. To bias or not to bias: that is the question. **Scand J Work Environ Health** 1983; 9: 57.
88. Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. **Biometrika** 1986; 75: 1–11.
89. Boivin J-F, Wacholder S. The case–cohort design for internal and external comparison. **Am J Epidemiol** 1985; 122: 522.
90. Tolonen N, Nurminen M, Hernberg S. Ten-year coronary mortality of workers exposed to carbon disulfide. **Scand J Work Environ Health** 1979; 5: 109–114.
91. Walker AM. Efficient assessment of confounder effects in matched follow-up studies. **Appl Stat** 1982; 31: 293–297.
92. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. **J Am Stat Assoc** 1983; 78: 1–12.
93. Hanley JA, Liddell FDK. Fitting relationships between exposure and Standardized Mortality Ratios. **J Occup Med** 1985; 27: 555–560.
94. Acheson ED, Gardner MJ. **Asbestos: The Control Limit for Asbestos.** London: H.M.S.O. for Health and Safety Commission; 1983.
95. Berry G. The mortality of workers certified by Pneumoconiosis Medical Panels as having asbestosis. **Br J Ind Med** 1981; 38: 130–137.
96. Liddell FDK. Excess PYLL for occupational mortality comparisons. **Int J Epidemiol** 1979; 8: 185–186.
97. Tsai SP, Wen CP. A review of methodological issues of the standardized mortality ratio (SMR) in occupational cohort studies. **Int J Epidemiol** 1986; 15: 8–21.
98. Liddell FDK. Statistical approaches to cohort studies in epidemiology: an overview. In: **Proceedings of the 45th Session, Amsterdam 1985: Invited Papers. Bull Int Stat Inst** LI, Book 1, Paper 3.1: 1–16, 1985.
99. McDonald JC. Health implications of environmental exposure to asbestos. **Environ Health Perspect** 1985; 62: 319–328.
100. Sorahan T. **Comparison of Expected Numbers using the Poisson Test of Statistical Significance: Power Considerations.** Birmingham: Cancer Epidemiology Research Unit, University of Birmingham; 1984 (privately circulated).

APPENDIX A

The Lexis Age–Period Space

In the age–period space described by Lexis [60], age (a) and date (d) are plotted on (cartesian) scales, vertical (usually downwards) and horizontal, respectively. All dates and ages are expressed in years (with decimals), and year has the same unit length on both scales. When a person's date of birth is denoted b, his age at date d is clearly $a(\text{at } d) = d - b$. Thus, his life-span can be represented by the straight line which bisects the angle between the horizontal axis and a vertical axis through b. Once a subject comes under observation, he remains "in view" until the end of follow-up, his earlier death, or his loss to the study. A subject's exposure may commence at any point and can continue to any other time, and may be of varying intensity or be interrupted on occasion; but cessation of exposure is *not* synonymous with loss to view. Note that, although the years when the subject is in view (i.e. under observation) are *judged* along the diagonal line, they are *measured* between verticals (from one date to another) or, equivalently, between horizontals (from one age to another).

Fig. A1 relates to the Quebec asbestos cohort [75], where birth before 1891 or after 1920 excluded from the cohort. For the Province of Quebec, mid-year populations and deaths from all causes (in 5-year age-groups, 15–84) were assembled for men for the years 1926 (when death registration became universal in Quebec) through 1975, and deaths by cause for 1951–75; the two rectangles on the diagram indicate where reference mortality was known. Within the second rectangle, 5-year squares have been delineated where relevant.

Five "case-histories" are illustrated, numbered (1) to (5) from left to right. Subject (1) was employed in 1905, at age 11, for 10 years; after a 7-year gap, he was employed again until 1947, and died in 1962 at the age of 68. Man (2) worked for 7 years (from age 12 until—probably—joining the army, in 1917), when all trace of him was lost. Person (3) had a single spell of employment, from 1923 for 15 years, and died while still employed. Man (4)—close to the "average" subject—started a 6-year spell of work at age 25 and survived beyond 1975. Subject (5) was employed 1942–61, survived into 1976, but died 3 years later. For all workers, the period of observation started with first employment and continued: either to end-1975 [for (4) and (5)]; or to death [for (1) and (3)]; or to loss to view [for (2)]. Note that (2) could not contribute to the analysis because his "loss" occurred outside the rectangles of known Quebec mortality. Also, the cause of death of (3) could not be taken into account, while the death of (5) was irrelevant—for the purposes of the study—because it occurred after the end of the follow-up.

Complex as such a diagram may appear, it has advantages in clarifying which subjects contribute useful information, in delineating what reference mortality rates will be useful, and in many other ways.

APPENDIX B

Comparing Subcohorts After Data Reduction by the Subject–Years Method

After subcohorts (k) have been defined by degree of exposure $[x]_k$, the data, for each age (a_i) and period (p_j), consist of: the number of events $(d_{ij})_k$ and the number of subject-years in view $(v_{ij})_k$. The exposure characteristics of cohort k is denoted $[x]_k$ to indicate it may be a vector of variables (such as duration and intensity of exposure to

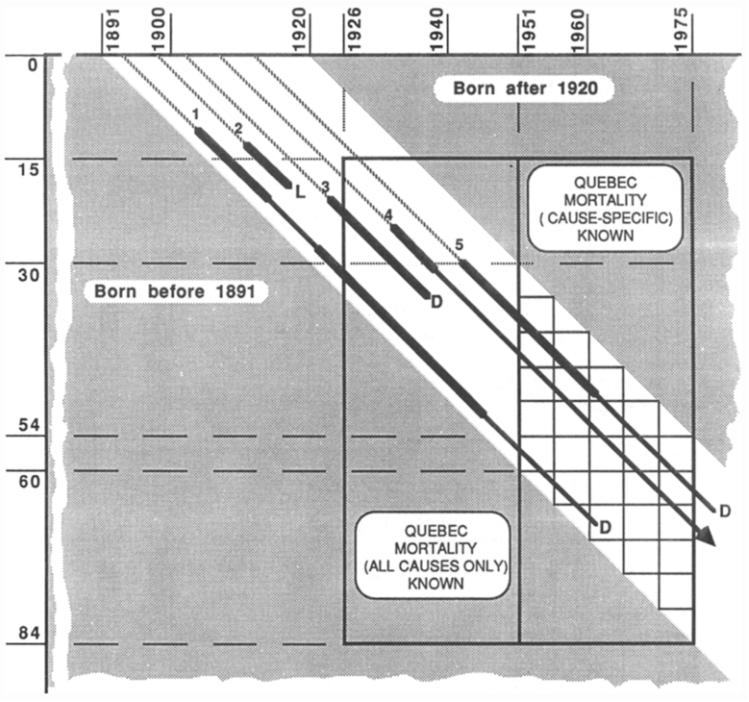

Fig. A1. The Lexis diagram for the Quebec asbestos cohort [75]. Abscissa: period (or year); ordinate: age (yr). The diagonal band represents the space through which the members of the cohort, born 1891 through 1920, were followed from first employment in the industry to the end of 1975 (or earlier loss or death). The lines 1–5 represent case histories: the thick bars indicate periods of employment; the broken lines periods before the particular subjects were employed (and so became "in view"); the continuous lines periods when the subjects were "in view". D = death of subject; L = loss of subject; arrow-head = subject known to be alive in 1976.

several potential hazards), or merely a single variable (such as average number of cigarettes smoked). The $(d_{ij})_k$ are often deaths due to a group of specific causes, but these are not differentiated in this Appendix; clearly, the same causes have to be assumed throughout. The $(v_{ij})_k$ are not strictly independent of the events, but can be assumed to be so [52].

In this Appendix, the fitting of any relationship between an outcome variable (say, y_k) for the kth cohort and any vector of explanatory variables (say, $[z]_k$)—which may also be considered as risk factors or covariates—is termed "fitting a model of y_k on $[z]_k$". Several types of model are displayed in Table B1; in types (1), (2), (3) and (6) the functional expression relating the outcome variable to the covariates would necessarily be complex, to allow for the form of relationship of the force of mortality with age. Further, in all types, interaction terms (such as cross-products of covariates) would probably be required.

Modelling without reference mortality experience

It would, in theory, be possible to fit a model of *type* (1). As the $(d_{ij})_k$ are subject to Poisson variation, the preferred method of fitting would be some form of Poisson regression [69]. Alternatively, one could calculate rates $(r_{ij})_k = \{(d_{ij})_k\}/\{(v_{ij})_k\}$ and fit a model of *type* (2)a. However, to gain the advantages of Poisson regression it would be appropriate to fit instead a model of *type* (2)b.

Another approach, perhaps oversimplifying, might be to assume it appropriate to pool over periods, by finding

$$(r^*{}_i)_k = \left\{\sum_j (d_{ij})_k\right\} \Big/ \left\{\sum_j (v_{ij})_k\right\},$$

and then to fit a Poisson model of *type* (3).

As the study of Doll and Hill [6] was over only a very short period of (less than 4.5 years), these authors could effectively ignore p_j and concern themselves with the (simplified) $(r_i)_k = \{(d_i)_k\}/\{(v_{ij})_k\}$. Then they could find a single summarized rate for each subcohort as

$$\bar{R}_k = \sum_i \{(\alpha_i)\cdot(r_i)_k\};$$

in practice the summary was obtained from standardization by the "direct" method, in which the α_i indicate the age distribution of some population and

$$\sum_i (\alpha_i) = 1.$$

They were then in a position to fit a model of *type* (4). This does not lend itself to Poisson regression, but in the particular circumstances of Ref. [6], where the lung cancer death rates were so closely associated with the amounts smoked in the four subcohorts (see Section 2.4), this was no serious drawback.

More generally, when p_j cannot be ignored, the corresponding approach would be to calculate

$$(\bar{R}_j)_k = \sum_i \{(\alpha_i)\cdot(r_{ij})_k\},$$

and fit a model of *type* (5) for each period separately. This would have the advantage over type (3) of allowing investigation of the consistency of the fitted models from one period to another. Nevertheless, there is the disadvantage that Poisson regression cannot be used.

Although model types (4) and (5) make use of an age distribution (α_i), it is well-known that *comparison* of standardized rates depends only marginally on the choice of the (α_i); indeed almost any other "reasonable" choice of weights would still allow satisfactory comparison between subcohorts.

Modelling with reference mortality experience

Much the more common approach has been to make use of reference mortality rates, one for each relevant cell in the Lexis space, and here termed m_{ij}. Then, for each cell, it is usual to find—for each subcohort separately—the terms $\{(v_{ij})_k\}\cdot(m_{ij}) = (e_{ij})_k$, which are usually called the "expected numbers of deaths, in subcohort k, within the cell for a_i and p_j".

It is possible to find the ratios $(\lambda_{ij})_k = \{(d_{ij})_k\}/\{(e_{ij})_k\}$, and, in theory, gain the advantages of Poisson regression by fitting a model of *type* 6. It is, however, more usual to simplify by obtaining a statistic which summarizes for each subcohort the values of $(\lambda_{ij})_k$; by far the most common is

$$\Lambda_k = \left\{\sum_{ij} (d_{ij})_k\right\} \Big/ \left\{\sum_{ij} (e_{ij})_k\right\},$$

which, because it is the average of the $(\lambda_{ij})_k$ weighted by the $(e_{ij})_k$, is equivalent [61] to a Standardized Mortality Ratio (SMR), standardization having taken place over both age and period. The usual notation is to term the numerator and denominator of Λ_k respectively O and E (subcohort specific, so here O_k and E_k), and to call Λ_k the O/E ratio for the kth subcohort, or its SMR. One can now fit a Poisson model of *type* (7) [93].

Models (6) and (7) require the age- and period-specific death rates (m_{ij}) in the chosen reference population. Such choice has often been found critical, particularly when—as too often—reliance is to be placed on the value of any single Λ_k, i.e. on the SMR for a particular subcohort.

Effects of simplification

It cannot be overemphasized that the seeming gains from the apparently simpler models, within the two types of modelling, may be spurious, by concealing variability between periods or over ages.

APPENDIX C

Power Justifications for Studies of Mesothelioma in Canadian Workers in Gas Mask Manufacture

In the McDonalds' study of mesothelioma after crocidolite exposure during gas mask manufacture [41], three

Table B1. Models for comparing subcohorts

Model	Outcome variable	Covariates: Degree of exposure	Age	Period	Subject-years in view
(1)	$(d_{ij})_k$	$[x]_k$	a_i	P_j	$(v_{ij})_k$
(2)a	$(r_{ij})_k$	$[x]_k$	a_i	p_j	—
(2)b	$(d_{ij})_k$	$\{(v_{ij})_k\}\cdot[x]_k$	$a_i\cdot(v_{ij})_k$	$P_j\cdot(v_{ij})_k$	—
(3)	$\Sigma(d_{ij})_k$	$\{\Sigma(v_{ij})_k\}\cdot[x]_k$	$a_i\cdot\Sigma(v_{ij})_k$	—	—
(4)	$\bar{R}_k$	$[x]_k$	—	—	—
(5)	$(\bar{R}_j)_k$	$[x]_k$	—	*	—
(6)	$(d_{ij})_k$	$\{(e_{ij})_k\}\cdot[x]_k$	$a_i\cdot(e_{ij})_k$	$P_j\cdot(e_{ij})_k$	—
(7)	O_k	$E_k\cdot[x]_k$	—	—	—

*For each period separately.

cohorts were involved: 113 workers (42% male) in a small factory at Asbestos, mainly incorporated within the Quebec asbestos cohort [75]; 32 workers (59% male) whom a foreman in a Montreal factory definitely remembered working on the gas mask filters; and 54 workers (80% male) at a plant in Ottawa who could be listed by a scientific officer (in whom a malignant pleural mesothelioma had been diagnosed in 1974) who was one of a small group "responsible for setting up and managing, at the time of the second world war, an assembly plant for gas masks for the Canadian army" [41]. The study of workers at the factory at Asbestos could be completed comparatively easily: all but one worker had been traced and 27 had died, four of them (15%) probably from mesothelioma. The background rate of mesothelioma, over many years, could be estimated as 2 per million population for both sexes [99], while in Canada annual deaths from all causes have numbered around 8000 and 6000 per million population of males and females, respectively. With a generous allowance for general environmental exposure, it is possible to infer a (sex-adjusted) expectation of 0.02 mesothelioma deaths at Asbestos among the 27 observed deaths; as four were from mesothelioma, the relative risk (RR) was around 200. With this information, were the investigations of the Montreal and Ottawa cohorts justified?

Assuming roughly the same proportions of deaths (all causes) among the 86 members of the two cohorts, and using the same methods of estimation, the expected number of mesothelioma deaths (for both cohorts combined) was 0.024. Because all employed at Asbestos were only assumed to have handled crocidolite, whereas those identified in the other two places were known to have done so, the RR in those places was not likely to be lower than at Asbestos. Even with true RR assumed as low as 125 (which corresponds to three cases), the power can be calculated, using the approach of Sorahan [100], as 95.0% for $\frac{1}{2}\alpha = 0.025$ and as 80.1% for values of $\frac{1}{2}\alpha$ down to at least 0.0005. The effort of tracing was clearly justified: in the event five cases were identified, putting the matter beyond reasonable doubt.