Soz Präventivmed 1992; 37: 13–21

0303-8408/92/010013-09 $1.50 + 0.20/0

Power calculation for cohort studies with improved estimation of expected numbers of deaths

Maria Blettner[1], Deborah Ashby[2]

[1] Institute of Epidemiology and Biometry, German Cancer Research Center, Heidelberg
[2] Department of Statistics and Computational Mathematics and Department of Public Health, University of Liverpool

Before starting any epidemiological or clinical study, it is essential to estimate, as accurately as possible, the power of the study to see whether the sample size is adequate to give a sufficiently large probability of rejecting the null hypothesis if there really is an association of interest. Methods and formulae have been discussed for different types of studies and under different assumptions[1–6].

Cohort studies in epidemiology start with the identification of a group of individuals for whom a certain exposure is measured. This group is then followed over time to collect data of either incidence or mortality of diseases for which it is hypothesized that the exposure may increase the risk. Two forms of analysis are conventional: internal and external comparisons. In the first, internal comparisons are made between individuals with the exposure and individuals without the exposure. In the second approach routinely collected data from external sources, e.g. national mortality rates, are used to estimate the number of expected deaths under the null hypothesis that the exposure has no effect; this is then compared with the observed number of deaths among the exposed individuals. The two methods have different strengths: the essential difference between internal and external comparisons is that for external comparisons the rate for the unexposed group is assumed known, whereas for internal comparisons it is estimated from the study population and therefore the estimation error has to be taken into account. Under some very general assumption it can be shown that external comparisons have a higher statistical power, but internal comparisons may be preferred to reduce biases such as the "healthy worker effect" in some occupational studies. This has been studied in detail[7,8].

Although these studies can be very expensive, sample size calculations seem to get less attention than in clinical trials, or in case control studies, as the total sample size is often constrained by the size of the cohort available in an occupational setting. However, given the resources required by cohort studies, two questions which should be considered are:

1. Is the study cohort sufficiently large to yield statistically reliable estimates of effects of interest?
2. Will the cohort have an adequate length of follow-up for studying delayed effects?[7]

In occupational cohort studies, the epidemiologist or statistician may wish to pose a different question, namely, what relative risk can be demonstrated from the cohort? Walter[1] took this approach and provided tables to find the minimum relative risk which can be detected given the sample size, the significance level and the desired power. In most practical situations it is not possible to increase the study size. The only alternative is to extend the follow-up time, which means waiting for a few more years until prolonged follow-up yields a worthwhile increase in the total number of person years and consequently in the number of expected cases for the disease of interest.

To calculate the power $(1-\beta)$ of a study one needs to decide the significance level (α), and the relative risk (under the alternative hypothesis) which the study should be able to detect. For the alternative hypothesis, the expected relative risk for an exposed group compared to an unexposed group depends both on the level of exposure and on the length of exposure, which will usually be unknown at the beginning of the study. The level of excess risk deemed to be important to detect is often to some extent arbitrary, but some prior knowledge from other studies may be available to estimate the magnitude of excess risk potentially attributable to the exposure of interest. The expected number of deaths is a function of the mortality rate applicable to the cohort, and the size of the cohort, their age and sex distribution and the length of follow up. Typically, the expected number of deaths is calculated naively but it is actually very sensitive to several assumptions implicit in this approach.

Sample size or power calculations can be made using either a hypothesis testing framework or a confidence interval approach. The power is then calculated as a function of the expected rates in the non-exposed and exposed groups, the sample size and of the significance level or specifications concerning the confidence intervals. Tables of the sample size requirement for cohort studies are available under very general assumptions[8].

Formulae are available for prospective studies when exposure is binary[2], categorical[4] or continuous[9] for calculating odds ratio or using logistic or log-

linear modeling. A review of these approaches and methods for more complex issues such as testing whether the gradient is equal to a specified non-zero value or testing for curvilinearity in the trend has recently been published [6].

Most of these papers discussing power or sample size calculations (whether using a hypothesis testing or a confidence interval estimation approach) have assumed that the number of expected deaths (E) in the study population is known. Alternatively, tables have been provided [1,3] for different values of E. None of the authors have emphasized that the actual estimation of E is an essential and critical step in any power calculation.

The present paper focuses on those assumptions which affect the estimation of the expected number for a particular study group. We concentrate on the hypothesis testing approach, although similar considerations apply if confidence limits are used, or even when the study is planned from a Bayesian perspective.

Any available information should be used in order to estimate the strength of the study. If adequate information is not available varying assumptions should be made to check the robustness of the sample size estimates.

This paper illustrates ways of calculating expected number of deaths for two different cohort studies with two different sets of *a priori* knowledge. It demonstrates how different assumptions about the latency (the time between the exposure and the onset of the disease) influence the expected number of deaths. It gives practical advice and discussion of two examples where estimation of sample size for prospective studies was required, paying particular attention to the calculation of the expected numbers under different assumptions about the age distribution in the population of interest. Even limited information should be used to improve the basis on which the power calculations are made in order to obtain more accurate estimates of the expected number of death. This problem appears to have received little attention in the literature and in practical planning of cohort studies.

Methods

We shall briefly review the calculation of the expected number (E) of deaths for a given cohort. The expected numbers of deaths are needed in order to calculate the power of a cohort study where tests for either the relative risk (RR) or the standardized mortality ratio (SMR) are performed.

The methodology for calculating expected deaths has been developed in the context of the analysis of cohort studies [10]. Various issues arising in their use are discussed in Gardner [11] but the principles are equally relevant for the design of the study and in particular for sample size calculation. The methods can be briefly summarized as follows: Each subject contributes to the calculation for those years in which she/he was at risk of dying. Let y_{ij} be the time in years that individual i spent in age group j and was at risk of dying of a tumor whose development may have been influenced by the exposure.

Denoting the annual mortality rate for age group j by d_j, the expected number of deaths can be expressed as

$$E = \sum_{j=1}^{G} \sum_{i=1}^{n} y_{ij} \cdot d_j \qquad (1)$$

where G is the number of age groups and n the total number of persons in the cohort. If d_j depends on the period of death, it can be allowed to vary.

We assume that the death rates are known from the appropriate reference population. Standard software for the calculation of the (y_{ij}) are available [12] and therefore the evaluation of (1) is straightforward.

Most papers on sample size calculations for cohort studies assume that the number of expected deaths is known or use the approximation $E = PY \cdot p_1$, where p_1 is the crude death rate of the general population. PY is estimated as the product of the number of persons in the cohort and the total time (in years) of follow-up. As can be seen from formula (1), E depends on the age composition of the cohort and the age specific mortality rates. In most situations the latter are known and the former have to be estimated.

Given the expected number of deaths (E), the probability of rejecting the null hypothesis if true (α), and the standardized mortality ratio (SMR) to be detected, the power of the study can be calculated assuming the observed number of deaths (O) follows a Poisson distribution. If the sample size is sufficiently large, a normal approximation is used. Correspondingly, if the comparision is made between two subgroups of the study population (e.g. unexposed versus exposed persons), the power calculation is based on the expected number of deaths in both groups (E_1 and E_2). Detailed formulae and some tabulations for different situations are given in Breslow and Day [8] (Formulae (7.1)–(7.4)).

Example 1

Cohort study in the automobile industry, UK

The first example demonstrates the calculation of the expected numbers of death for a cohort where some a priori information on the age distribution and dates of entry to the cohort was available.

Study design:

Merseyside is a major industrial conurbation with high rates of many cancers. One hypothesized agent

for lung cancer is oilmist exposure, which may raise an individual's risk by about one third[13]. An epidemiological study was proposed to compare the mortality of approximately 10,000 exposed workers in two automobile factories in Merseyside both with that expected on the basis of national rates, and with controls from the same factories.

Initial estimate of expected deaths

A crude calculation of the expected number of deaths gives an estimate of 50 lung cancer deaths which would be expected from a cohort of 10,000 men, assuming an overall lung cancer mortality rate of 1/1000/year[14] and an average follow-up time of 5 years. Applying formula (7.1) from Breslow and Day[8] yields a power of 60% for a relative risk of 1.34 and a power of 90% for a relative risk of 1.5 at a 5% (two sided) level of significance. Assuming an average follow-up time of 10 years doubles the expected number of cases and increases the power to 88% and 99.4% respectively. Similarly, assuming an average follow-up time of 20 years redoubles the expected number of cases to 200 and increases the power to more than 99% in both cases.

If one wishes to make assumptions about latency the follow-up time simply starts with time appropriately lagged. For example, a latency of ten years with a further follow-up of five years yields the same estimate of the expected number as assuming no latency and just five years follow-up. These calculations take no account of the ageing of the cohort, and this is addressed in the next section.

Estimation of expected death using prior information

In the feasibility study for a cohort study which was proposed in two automobile factories in the UK, one of the companies provided a list of their employees which included identification numbers, date of birth, and date of entry into the company for about 4700 men. Everybody on this list had worked in the area of interest for the planned study, but the list included only those who started work between 1963 and 1978, whilst the study period includes the year 1963 until 1988. Nevertheless, the available data enabled detailed sample size calculation. Discussion with personnel managers indicated that the list could be regarded as a good approximation to the whole study group, as the number of people who started work after 1978 was small. Also, employees who started work after 1978 would not contribute many person years to the age groups of interest in the first analysis, as they would still be relatively young in 1988.

This list was used to estimate the age distribution of the cohort and to calculate the number of deaths expected from different cancer sites. Table 1 shows the number of men who started work in the company by calendar year and by 5 year birth intervals. The table shows that the majority of the population started work prior to 1969 and that in recent years fewer middle aged men were employed. The year of birth was unknown for 5% of the population.

The total number of person years and expected number of deaths were calculated under the following assumptions:

1. In order to determine the distribution of different age groups (Table 1) it was assumed that everybody was followed up until 1.1.1988. Correction of the risk set for deaths from other causes were made applying the corresponding U.K mortality rates[14] with the assumption that all death occur at the mid point of the ten years age group. Although this assumption is violated for older age groups, it is taken as a reasonable first approximation.

Tab. 1. Number of persons by year of birth and year of employment.

Year of birth	1910–19	1920–29	1930–39	1940–49	1950–59
Age[1]	73	63	53	43	33
Year of employment					
1963	4	88	118	25	
1964	26	222	388	119	
1965	16	159	271	94	
1966	8	101	154	107	
1967	3	59	160	171	
1968	2	160	300	360	
1969	2	98	180	226	9
1970/71		24	50	57	49
1972		6	16	19	3
1973		31	63	104	38
1974/75		13	25	55	12
1976			26	62	10
1977			13	31	7
1978			20	33	55

[1] Age is mean age in 1988 for each birth cohort.

Tab. 2. Death rate per 1 000 000 males for selected causes by age at death.

Age Group	All causes	All cancer	Lung cancer	Stomach cancer	Pancreatic cancer	Large intestine and Rectum
25–35	1,032	182	16	7	4	13
35–44	2,114	480	119	41	20	57
45–54	7,125	1873	778	182	88	196
55–64	19,627	5753	2690	613	257	557
65–74	53,749	13787	6049	1582	597	1509
75+	139,999	22263	6750	2582	968	3303

(Source: ref 14).

Tab. 3. Expected number of death in a cohort of 4700 men.

Age Group	Person Years	All causes	All cancer	Lung cancer	Stomach cancer	Pancreatic cancer	Large intestine and Rectum
a) 0-yrs latency							
<24	2150						
25–34	23,200	23.9	4.2	0.4	0.2	0.1	0.3
35–44	35,500	75.0	17.1	4.0	1.5	0.7	2.0
45–54	24,000	171.0	45.0	18.7	4.4	2.1	4.7
55–64	8,000	157.0	46.0	21.5	4.9	2.2	4.5
65–74	580	31.2	8.0	3.5	0.9	0.3	0.9
75+	7	1.0	0.2	0.0	0.0	0.0	0.0
Total	91,300	459.1	120.4	48.3	11.8	5.4	12.4
b) 10 years latency							
25–34	2,000	2.1	0.4	0.0	0.0	0.0	0.0
35–45	22,000	46.5	10.6	2.6	0.9	0.4	1.3
45–54	18,000	128.3	33.7	14.0	3.3	1.6	3.5
55–64	8,000	157.0	46.0	21.5	4.9	2.1	4.5
65–74	580	31.2	8.0	3.5	0.9	0.3	0.9
75+	10	1.0	0.2	0.0	0.0	0.0	0.0
Total	50,600	366.1	98.8	41.7	10.0	4.4	10.2
c) 20 years latency							
25–34	400	0.4	0.1	0.0	0.0	0.0	0.0
35–44	2300	4.9	1.1	0.3	0.1	0.0	0.1
45–54	6500	46.3	12.2	5.1	1.2	0.6	1.3
55–64	4500	88.3	25.9	12.1	2.8	1.2	2.5
65–74	480	25.8	6.6	2.9	0.8	0.3	0.7
75+	10	1.0	0.2	0.0	0.0	0.0	0.0
Total	14,200	166.7	46.0	20.4	4.8	2.1	4.7

2. Death rates for 10 years age groups were taken from rates published for the U.K. for the years 1968–1978 (see Table 2). It was assumed that the rates were valid for the period from 1963 to 1988.
3. Three different assumptions about the latency were made to calculate the person years: firstly, immediate effect of the exposure (i.e. no latency period), secondly, a 10 year latency period and lastly a 20 year latency period.

Table 3 was calculated under each of the assumptions about the latency. Under the first of these assumptions, the total number of person years in the population of the 4700 men is 91,000, which corresponds an average follow-up time of nearly 20 years. This can be explained by the fact that most men were employed around 1965 and were in their thirties, so that most of them were still alive in 1988. With the second and the third assumption the total number of person years of interest drops to 50,000 or 14,000 respectively, because the early years of exposure are eliminated from the analysis. This clearly demonstrates the importance of taking into account the appropriate assumptions about the latency periods in estimation the person years of the cohort.

Expected numbers

Expected numbers were calculated by applying (1) to all causes of death combined, all cancers combined, lung cancer, stomach cancer, pancreatic cancer and cancer of the large intestine and rectum combined. These locations have been chosen as elevated relative risks have been reported[15]. The

results are presented in Table 3a, 3b and 3c. In Table 3a the total number of expected deaths is about 460 of which 120 cancer deaths are expected, including 48 lung cancer deaths and 12 stomach cancer deaths. Comparing Tables 3a and 3b shows that the total number of expected deaths differs by about 100 between the first and second assumptions. However, the effect on the total number of cancers, mainly at the sites of interest (lung, stomach) is less critical although proportionately the same. If we assume a latency period of 20 years, the number of expected cancer cases drops to 46 in total and only 20 lung cancer cases will be expected. The small difference between the expected cancer cases in Table 3a and Table 3b can be explained by the fact that most persons in the cohort were employed in the 1970's and at a very young age. Their contribution to the expected number of death is low between 1970 and 1980, but becomes more important thereafter.

No detailed information was available for the second company. However, after discussion with personnel managers and the medical officer, it was be assumed that the situation was similar, i.e. the companies started in the same year and production declined at the same time. The second company, however, had about 6000 employees in the area of interest. It seems justified to perform the power calculation for the two-factory study on the basis of a total work force of 10,000 workers with the age distribution as given in Table 1. Under the null hypothesis of no excess risk due to exposure, this assumption yields a total of 100 expected lung cancer deaths.

Power calculation

Table 4a shows the power of the study if the mortality experience of the cohort is compared to mortality rates from the general population for different numbers of expected cases and different SMRs. We are assuming that everybody in the cohort is exposed.

It can be seen that the study will have sufficient power to detect SMRs of the order of 2 for those cancers which are not too rare, such as lung and stomach cancer. The study will not be sufficiently large to detect even a twofold excess risk in pancreatic cancer. Under the assumption that the latency time is ten years. 80–90 lung cancer cases are expected, giving a power of 80% to detect any SMR higher than 1.3. However this is only true for the total population. Subgroup analysis will be less powerful and will only detect high excess risk. If a latency period of more than 20 years has to be assumed the power of the study is much smaller. Only 40 of the expected 100 lung cancer deaths will occur in a time period that starts 10 years after the exposure. However, this will still be sufficient to detect a SMR as low as 1.5 with a power of 86%.

Tab. 4. Power, given RR or SMR, α and expected number of cases.

a) External Comparison.

	Expected number of cases				
SMR	80	60	40	20	10
1.20	50	42	30	18	15
1.30	80	69	52	30	22
1.40	90	88	72	45	31
1.50	95	96	86	60	43
2.00	100	100	100	97	84
3.00	100	100	100	100	99

Probability of declaring significant ($\alpha = 0.05$, one-sided) difference.

b) Internal Comparison, equal size groups.

	Expected number of cases in each group				
RR	80	60	40	20	10
1.20	33	27	21	15	11
1.30	55	45	35	22	15
1.40	75	64	50	31	20
1.50	89	79	64	41	26
2.00	100	100	98	84	58
3.00	100	100	100	100	96

Probability of declaring significant ($\alpha = 0.05$, one-sided) difference.

For an internal control group with a similar age distribution as above and the same sample size (about 10,000 men), we expect the same number of cases in the control group. Given the significance level of 0.05, we are able to detect a relative risk of 1.5 with a power of 85%. The power of the study for different combination of expected numbers of cases and relative risks is given in Table 4b.

Summary and comments for Example 1

The above calculations clearly demonstrate the weaknesses and lack of power of the planned study. The crude calculation yields an expected number of lung cancer deaths of 200 after 20 years follow-up, while the more accurate estimates shows that the expected number is only 80 if a latency time of 10 years is taken into account. The over estimation of the power of the study can be explained by the crude assumptions about the death rates, and in particular not taking into account the young age of the cohort at the beginning of the study period. Also the estimation of the total follow-up time was too optimistic. It also shows the importance of further follow-up for about 10 years. The expected numbers for stomach cancer and other cancers of the digestive system will then reach a level at which the power of the study would be sufficient to detect relative risks in the order of 2, which have been reported in the literature[11].

Example 2

A prospective cohort study among workers in nuclear power stations, FRG

Power calculations are particularly relevant to the study of cancer or death among workers in nuclear power stations since uncommon events and a small population lead to small expected numbers and low power. In general, the sensitivity of a statistical test can be enhanced by increasing the expected number. However in this particular case, as in many occupational cohort studies, the population size cannot be enlarged. If the study population is enlarged by unaffected persons (employees working in the nuclear power plant but not being monitored) any effect would be diluted. Pooling data from several plants is necessary, and pooling data from different countries is a possible way to gain power. However grouping must be performed with care as different exposure regulations exist in different countries.

In order to assess the feasibility of a prospective cohort study among workers in nuclear power stations in West Germany, information was gathered about the work force in the plants. Production of electricity from nuclear power started in the Federal Republic of Germany in 1961. Since then, 27 reactors have started production. It is known that between 1967 and 1985 approximately 4000 persons worked in the plants and it has been estimated that approximately 50,000 person years were accumulated for plant workers (see Table 5). Information on the collective dose for this period is available and is about 200 person Sv (Sievers). As no exact information on the age distribution of the cohort could be obtained prior to the study, power calculations were carried out with several assumption about the age distribution of the employees. In contrast to the first example, information on the exposure was available and could be used to estimate the relative risk associated with the exposure.

Federal regulations require that all persons who are working in nuclear plants are registered in a central registration office, so the total number of employees was available for each year. The number of people who started employment in a certain year was calculated by taking the difference between two adjacent years. This approximation is justified as few people have left the industry. Information was available on the time each plant started to produce electricity. Major increases in the total number of surveyed persons correspond to the years when new plants were opened. No information on the exact age at entry in the company could be obtained. Two different assumptions were made: a) the average age at entry is about 30 years, implying that the cohort is still very young, b) average age of 35 years which was the best "expert" estimate from medical officers in the plants. Table 5 shows the number of people by year of employment since 1967, and accumulated person years until 1991. As the cohort is young, it can be assumed that the loss due to death is small. However as the follow-up system in West-Germany is complicated, 5% of the total person years were deducted to allow for the loss to follow-up. This estimate was based on experience in other follow-up studies in West-Germany.

Tab. 5. Number of persons by year of employment and estimated person years until 1991.

Latency Group	Year of first employment	Number of persons	Person-years
> 20 yrs	1967	102	2448
	1968	7	161
	1969	240	5280
	1970	5	105
	1971	144	2880
Total		498	10874
15–20 yrs	1972	188	3572
	1973	64	1152
	1974	340	5780
	1975	75	1240
	1976	412	6180
Total		1020	17924
10–15 yrs	1977	421	5894
	1978	0	0
	1979	0	0
	1980	225	2475
	1981	344	3440
Total		990	11809
5–10 yrs	1982	183	1647
	1983	486	3888
	1984	595	4165
	1985	196	1176
	1986	100	500
Total		1560	10376
Total		4068	50983

Mortality rates for the general population for males were used to calculate the expected number of deaths for leukemia and lung cancer. The occurrence of leukemia should be the most sensitive indicator of a radiation effect since it is the disease that has been most closely linked with radiation in other studies and it has a relatively short latent period. Lung cancer has also been linked with radiation in occupational studies [16] and lung cancer deaths are likely to be the most common in this cohort.

Table 6 shows the expected number of lung cancers for different age groups and in total, taking into account a latent period of five years and assuming that the workers were, on average, 30 years old when they started employment in the nuclear power plants. The bottom line of Table 6 gives the corresponding numbers under the second assumption (age 35 years) and demonstrates clearly the importance of the assumptions about age-distribution. Table 6 also emphasises that a longer

Tab. 6. Expected number of lung cancer death in the cohort. Age at start: 30 years, Latent period: 5 yrs.

Age Group	1991	1996	2001
60–64		2.00	5.25
55–59	8.36	9.00	12.70
50–54	3.90	5.40	6.90
45–49	1.12	2.20	3.40
35–39	0.10	0.30	0.30
Total	13.48	18.90	28.55
Total[1]	27.80	37.90	56.50

[1] if age at start is 35 years in average.

Tab. 7. Expected number of leukemia death in the cohort. Age at start: 30 years, Latent period: 5 yrs.

Age Group	1991	1996	2001
60–64		0.2	0.4
55–59	0.6	0.7	0.9
50–54	0.5	0.8	1.0
45–49	0.2	0.3	0.6
35–39	0.0	0.1	0.1
Total	1.3	2.1	3.0
Total[1]	2.15	3.15	4.62

[1] if age at start is 35 years in average.

follow-up (until 1996 or 2001) will increase the number of expected deaths substantially. Not only will more person years have been accumulated, but there will also be more people in older age groups which have higher mortality rates.
Table 6, together with the power calculations in Table 4, shows that the power to detect a relative risk of 1.5 is 60%, assuming that the average age at starting was 30 years, while the power increases to 86% under the assumption that the men were 35 years when entering the plants. Table 7 gives corresponding numbers for leukemia. As leukemia is a very rare disease only a few cases would be expected during the next few years.

Summary and comments for Example 2

Risk estimates and models developed from investigations in populations which have been exposed to high-dose radiation can be used to estimate the risk of persons exposed to low-dose radiation. The main purpose of the cohort study among employees in nuclear power stations is to determine the extent to which direct observation confirm or contradict these estimate and models. For lung cancer, the induction rate has been reported to be between $0.0003\ Sv^{-1} \cdot a^{-1}$ and $0.0009\ Sv^{-1} \cdot a^{-1}$ (0.0003 per Sv per year). We can assume that the life time dose among workers in nuclear power stations is no higher than 5 rem (50 mSv), which yields a relative risk of less then 1.1 for lung cancer. The above calculation demonstrates that a power sufficient to detect such a small risk will not be achieved during the initial ten years of follow-up.
For a relative risk of less than 1.1, we obtain a power of less than 20% taking all lung cancer cases up to the year 2001. We have to conclude that the study will not be able to detect a risk which would be estimated from current risk coefficients. The above calculation alone would not support the decision to conduct the proposed study as the sample size calculation shows clearly that the study will not have enough power to detect any reasonable risk. However, the conduct of this study is not guided solely by power considerations.

Discussion

The paper has discussed some practical aspects of sample size calculations in cohort studies and summarized the basic methods which have been discussed in the literature. Our main goal was to highlight the assumptions which need to be made, especially regarding the expected number of deaths, prior to performing power calculations. We have used the conventional approach which focusses on significance tests to demonstrate the effect of different assumptions on the power, but identical considerations would apply if a confidence interval or Poisson regression approach were considered.
Given the expected number of deaths the power calculation is straight forward. Different approaches have recently been reviewed [6].
The expected number of deaths is dependent not only on the total size of the cohort, but also on its age-sex structure together with assumptions about latency and minimum period of employment needed for an increased risk. Only the exposed part of the cohort contributes to the question of interest. It is not sufficient to know the total number of employees. Moreover, if it is appropriate to take a latency period into account, a large number of the accumulated person years are of no relevance for the study as cancers occurring within this latency period could not be attributed to the exposure. A distinction should be made between latency period and minimum time of exposure of interest. In occupational cohort studies it is sometimes recommended that data only be collected on people who have been exposed for at least, say, two or five years. This may decrease the size of the population of interest substantially. Additionally, competing risks may decrease the population at risk, particularly if the diseases of interest are rare and the population is at a high risk of dying from other causes. The calculation should allow also for loss to follow-up, which in countries without a good working registration system can be substantial.

For the death rates the following issues are important. Firstly, the use of age and sex-specific death rates is necessary as for most diseases the death rates increase sharply with age and vary between the male and female population. Secondly, the decision has to be taken as to which are the appropriate death rates for external comparisons. In general, using national death rates seems to be adequate for the power calculations, but it should be borne in mind that differences due to regional variation or inappropriate comparison groups ("healthy worker effect") may lower the power considerably.

When a cohort study is planned, there are many questions which should be addressed to the organization at a very early stage. Much of the required information could be obtained by intensive discussions with the personnel managers, workers of the company or by looking at the company's history. As we have demonstrated in our second example some very basic assumptions about the population may help to avoid large errors and a huge over estimation of the possible power of a study.

If data are available before the start of the study, then further calculations are possible to evaluate whether and after what time period the data can usefully be analysed. This strategy would help to avoid uninformative interim analysis and identify the appropriate follow-up time to assure that the power will be sufficient for analyzing particular questions of interest.

In our approach we have discussed how to estimate the expected numbers under the null hypothesis. We have assumed that reasonable estimates or prior information on the magnitude of the relative risk are available, either from other studies or from more general biological considerations. To estimate the relative risk associated with the exposure it is important to have some prior knowledge about the level of exposure and the length of exposure. If this is not available, the concept of "minimum detectable relative risk" is very attractive and could be applied to evaluate the importance of a planned study.

Summary

This paper focuses on improving the accuracy of sample size calculations for cohort studies by careful calculation of the expected number of deaths in the population, taking into account either prior information or realistic assumptions about variables which may affect the mortality or incidence. Sometimes small changes in the assumptions can dramatically alter the expected numbers and may necessitate modifications in the design of the study. Possible modification include extension of the follow-up time, and recognition that the real strength of the study may lie in the potential for pooling several similar studies. The problem will be discussed with reference to two examples of occupational cohort studies where differing prior information was available.

Résumé

Calcul de puissance pour les études de cohortes, avec amélioration de l'estimation du nombre attendu de décès

Cet article concerne la précision des estimations de taille d'échantillons pour les études de cohortes. Le calcul précis du nombre de décès attendus dans la population prend en compte les variables susceptibles d'affecter la mortalité ou l'incidence, provenant soit d'une connaissance préalable, soit d'hypothèses réalistes. De modestes changements d'hypothèses peuvent parfois altérer de façon substantielle les nombres attendus et nécessiter des modifications dans le protocole de l'étude. Parmi les modifications possibles, il faut citer la prolongation du temps de suivi de l'étude ainsi que le constat que la valeur réelle de l'étude pourrait reposer sur la possibilité de mise en commun de plusieurs études similaires. Le problème est discuté à l'aide de deux exemples d'études de cohortes professionnelles pour lesquelles différentes informations préalables sont disponibles.

Zusammenfassung

Berechnung der statistischen Macht in Kohortenstudien mit verbesserten Schätzungen für die erwartete Zahl der Todesfälle

Diese Arbeit beschäftigt sich mit der Genauigkeit der Berechnung des Stichprobenumfangs in Kohortenstudien, wenn detaillierte Berechnungen für die erwartete Zahl der Verstorbenen berücksichtigt werden. Dies kann entweder durch die Ausnutzung vorhandener Informationen oder durch realistische Annahmen über die Faktoren, die Mortalität oder Inzidenz beeinflussen, geschehen. Schon kleine Unterschiede in diesen Annahmen kann die erwartete Zahl der Verstorbenen erheblich verändern und es notwendig machen, das Design einer Studie zu verändern. Solche Modifikationen bestehen z. B. in der Verlängerung der Follow-up Zeit der Studie oder in der Einsicht, dass es nötig ist, Daten aus mehreren Studien zusammenzufassen. Die Probleme werden anhand von zwei Beispielen aus dem Bereich der Berufsepidemiologie diskutiert.

Acknowledgement

The study in the Merseyside automobile industry has been funded by a grant from the North West Cancer Research Fund.

References

1 *Walter SD.* Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. Am J Epidemiol 1977; *105*: 387–397.
2 *Beaumont JJ, Breslow NE.* Power considerations in epidemiologic studies of vinyl chloride workers. Am J Epidemiol 1981; *114*: 725–734.
3 *Gordon I.* Sample size estimation in occupational mortality studies with use of confidence interval theory. Am J Epidemiol 1987; *123*: 158–162.
4 *Howe GR, Chiarelli AM.* Methodological issues in cohort studies II: Power calculations. Int J Epidemiol 1988; *17*: 464–468.
5 *Phillips AN, Pocock SJ.* Sample size requirements for prospective studies, with examples for coronary heart disease. J Clin Epidemiol 1989; *42*: 639–648.
6 *Lubin JHY, Gail MH.* On power and sample size for studying features of the relative odds of diseases. Am J Epidemiol 1990; *131*: 552–566.
7 *Checkoway H, Pearce NE, Crawford-Brown DJ.* Research methods in occupational epidemiology. Oxford University Press, 1989.
8 *Breslow NE, Day NE.* Volume II – the design and analysis of cohort studies. Statistical methods in cancer research. Lyon: IARC Scientific Publications, No. 82, 1987.
9 *Whittemore AS.* Sample size for logistic regression with small response probability. Journ Amer Stat Ass 1981; *76*: 27–32.
10 *Berry G.* The analysis of mortality by the subject-years method. Biometrics 1983; *39*: 173–184.
11 *Gardner MJ.* Expected numbers in cohort studies. Scientific Report No. 6, Medical Research Council Environmental Epidemiological Unit. 1984.
12 *Coleman M, Douglas A, Hermon C, Peto J.* Cohort study analysis with a FORTRAN computer program. Int J Epidemiol 1986; *15*: 134–137.
13 *Alderson MR.* Occupational cancer. London; Butterworths, 1986.
14 *Gardner MJ, Winter PD, Taylor CP,* Acheson ED. Atlas of cancer mortality in England and Wales, 1968–1978. Chichester: Wiley, 1983.
15 *Silverstein M, Park R, Marmor M, Maizlish N, Mirer F.* Mortality among bearing plant workers exposed to metalworking fluids and abrasive. J Occup Med 1988; *30*: 706–714.
16 Advisory Committee on the Biological Effects of Ionizing Radiation (BEIR III). National Academy of Sciences-National Research Council. The Effects on Populations of Exposure to Low Levels of Ionizing Radiation. Washington, D.C.: U.S. Government Printing Office. 1980.

Address for correspondence:
Dr. Maria Blettner
Institute of Epidemiology and Biometry
German Cancer Research Center
Im Neuenheimer Feld 280
D-6900 Heidelberg, Germany