Design and Planning of Epidemiological Studies

I.12

Pascal Wild

12.1 Introduction 465
12.2 Early Planning 465

Objectives – the Concept of the Study 465
Scientific Background 466
The Study Protocol 466

12.3 Design 469

Study Base 470
Cohort Studies 470
Case-Control Studies 474
The Study Base Principle: Selection, Exclusion and Resulting Bias 478
Choosing Between Epidemiological Designs 479
Statistical Power 481

12.4 Measures of Disease Outcome and Exposure Parameters 482

Measurement and Classification of Exposure 482
Measurement and Classification of Health Outcomes 487
Measurement and Classification of Health Outcomes 488
Information Bias 488

12.5 Confounding 489
12.6 Statistical Analysis 490
12.7 Practical Issues 491

Fund Raising 491
Data Management 492
Quality Assurance 493
Study Conduct: Manual of Operations 494

Time Line 495
Project Diary 495
Ethical Aspects 497
Scientific Collaborations and Multi-centre Studies 497
Publication 498

12.8 Conclusions 499
References 499

Introduction 12.1

This chapter deals with practical issues in designing and planning analytical epidemiological studies. Although most of the practical issues are consequences of the theoretical principles of epidemiology as presented in standard textbooks (Breslow and Day 1981, 1987; Rothman and Greenland 1998; Miettinen 1985) and in Chaps. I.1, I.5, I.6, I.9, II.1, II.5, and II.6 of this handbook, the emphasis here is on how to proceed practically when planning a study.

This chapter is based on our experience in conducting epidemiological studies and on a series of references in which many of the concerns in the practical planning of studies have been described at length. Among the key sources we used are the books by Hernberg (1992) and Armstrong et al. (1994). The series of papers by Wacholder et al. (1992a,b,c) have also inspired much of our writing. It will start with a section on early planning in which the general setting of any study is described as well as the key planning document, i.e. the study protocol. The second section is devoted to the choice and implementation of an actual design where we focus on cohort and case-control studies. The next section focuses on data collection, both, with respect to the exposure and the disease outcome. The final section is devoted to practical issues and gives a list of topics arising in all studies that may not always get the appropriate attention while planning an epidemiological study.

Early Planning 12.2

Objectives – the Concept of the Study 12.2.1

The first step in the planning of an epidemiological study is the definition of the problem. Researchers must ensure that they have a clear view of the problem at the abstract-general level. At this conceptual level a problem takes the form "does X cause Y?" or "how much will a certain amount of exposure to X affect Y?" (Hernberg 1992, Chap. 4). It should be stressed that in analytical studies, researchers are interested in the relationship between exposure and disease which would be valid in other circumstances. They are only interested in the particular morbidity experience of the study population as far as it can be extrapolated to other populations.

The general interest of the investigator first has to be translated into precisely formulated, written objectives. A limited number of study objectives should be defined. These objectives may be of two kinds.

A first case is when the study is focused on specific analytical questions with a predefined hypothesis. For instance "does exposure to extremely low frequency (ELF) electromagnetic fields cause childhood leukemia?". Here the hypothesis should be formulated as a series of operational questions. One can specify the ELF fields in a variety of ways both qualitative (yes/no or low/medium/high) or

quantitative (e.g. present intensity, mean intensity over the last years or cumulative exposure). Other pre-defined operational hypotheses may include subgroup analyses based on the disease subtype. These operational hypotheses will have to be clarified by confirmatory tests. Additional results based on the observed data will then be clearly identified as such.

A second study type focuses on broader hypotheses generating questions that will be investigated by exploratory analyses such as "what occupations are associated with an increased risk of laryngeal cancer?". Even in this case a predefined list of questions is useful. In the present example, this would consist in a list of occupations considered. Again an unsuspected excess in an occupation not considered a priori should be identified as such.

12.2.2 Scientific Background

It is of crucial importance to undertake a thorough literature search and to know the literature in detail before planning any new project. Occasionally, the literature review may show that the answer to the study question is already available and that further data collection is not needed.

Evaluating epidemiological evidence from the literature is often challenging even for experienced researchers. Because a single positive finding may be a chance finding, a complete literature search, including negative results, should be conducted. One should consider systematic errors (biases) and confounders that may have led to a particular result in previous studies. Several independent studies using the same design and the same procedure of data collection may have similar results due to common biases or confounding. It is therefore important that the sources of spurious results are identified and controlled in subsequent investigations. It would be ineffective to simply replicate previous studies, without consideration of new research questions raised by previous studies, that could not be addressed because the information was not collected. A large literature body on how to perform a systematic literature synthesis is available, see Chap. II.7 of this handbook and references therein.

The literature search should, however, not be restricted to epidemiological studies, but may encompass a large range of topics from biological mechanisms or biomarkers related to the hypothesis under study to techniques of exposure measurement.

12.2.3 The Study Protocol

An epidemiological study is generally a complex undertaking of long duration, requiring time from study investigators and technicians, large resources in personnel and funding. The success of a study depends on a careful preparation. It is self-evident that such an undertaking cannot be done without a written study protocol.

A study protocol (study plan) should cover all aspects of the planned study. It should first state the precise objective of the study and describe the scientific

rationale for undertaking a new study, based on a literature search of the relevant publications (scientific rationale: study background and objectives). It should then define the study design including the precise study base and an estimate of the corresponding statistical power to achieve the objectives. Finally it should give a detailed account of what the epidemiologists intend to do and how they intend to do it. This entails for instance procedures for identifying the outcome parameter, measuring of exposure and confounders, data management and all steps taken for quality assurance. These aspects could be contained in a separate operations manual. It should also address the strategy for statistical analysis, ethical considerations and data protection procedures, project organization, quality control, time schedule and study diary, publication, and budget.

According to Miettinen (1985) a study protocol should have five purposes:

- Crystallize the project to the researchers themselves
- Give referees the possibility to review the project (especially for funding)
- Inform and educate all those taking part in the project
- Ensure the main researchers do not forget any details of the plan in the course of the study
- Document the procedures of the project for the future.

In summary, a protocol must be so detailed that an independent researcher could carry out the study based on it. The planning of each epidemiological study needs explicit and operationalized hypotheses that have to be formulated as specific and precise as possible. The selection of the population under study has to be justified in light of the research question.

An outline of the issues to be covered in the study protocol is given in Box 1. Most of these issues will be discussed in what follows. For further details we refer to Chap. I.13 of this handbook.

Box 1. Overview of a study plan and corresponding key problems to be addressed in the planning phase

1. *Research question and working hypotheses*
   - Relevance/previous findings
   - Choice of an appropriate target population
   - Problem: Operationalizing, i.e translation of items into variables that can be quantified
   - Precise definition of endpoint
   - Precise definition of independent variables and confounders
   - Choice of statistical measures (proportions, means, risks)
   - Translation into a hypothesis that can be tested statistically
   - Confirmatory testing of hypothesis/exploratory analyses

box to be continued

2. *Study design*
   - Optimal design (theoretical)
   - Practical limitations/feasibility

3. *Study base (target population) & study population*
   - Often limited access to subjects
   - Problem: Generalization from study sample to target population

4. *Size of study & its justification*
   - Sample size determination depends on precise definition of hypothesis
   - Size of acceptable type II error to be considered (power!)
   - Often: Power calculations based on fixed sample size

5. *Selection and recruitment of study subjects*
   - Problem: Selection bias (survival bias, referral bias)
   - How to assure representativeness
   - Means to maximize response
   - Sampling procedure/
   - Data source
   - Potential problem: Finding appropriate reference groups
   - Matching, if applicable

6. *Definition of procedures for measurement & data collection of variables*
   - Problem: Information bias (recall bias, measurement bias)
   - Instruments (postal, face-to-face, telephone)
   - Structure? Comprehensible? Answerable? Length?
   - Sensitive issues?
   - "Objective" sources of data? Measurements?
   - Guidelines for measurements
   - Usefulness in the analysis
   - Coding

7. *Exposures/risk factors/potential confounders & effect modifiers*
   - Problem: Assess confounders!
   - Precise definition
   - Valid assessment
   - Quantification: What information is needed?
   - Chronological reference

box to be continued

8. *Concept for data entry and data storage*
   - Data base
   - Ergonomic layout of data entry screen
   - Validation of data entry
   - Validation of coding
   - Plausibility checks/data corrections (concurrent to data collection)
   - Documentation: Always keep raw data!
   - Merging of data

9. *Strategy for analysis including statistical models*
   - Compliance with a priori hypotheses
   - Mark and report ad hoc hypotheses accordingly
   - Problem: Adequate consideration of confounders

10. *Measures for quality assurance*
    - Guidelines for data collection, measurement, interview (interviewer and operations manual)
    - Training of staff
    - Minimal information on non-responders
    - Validation of data
    - Comparison with reference data external to the study
    - Description of changes of original study plan
    - Project diary

11. *Measures to guarantee confidentiality & ethical principles*
    - Obtain informed consent by participants
    - Anonymize data, keep names separate

12. *Time line & responsibilities*
    - Chronological sequence
    - Scientific councils
    - Agreement on publication rules

12.3 Design

The study design governs all procedures for selecting and recruiting individuals in the study sample. A design may be chosen depending on the study objectives, but may also rely on practical issues such as costs, or data availability. Most epidemiological studies, with the exception of clinical trials and intervention studies, are purely observational, in that the investigator cannot assign the exposure to the study subjects as in experimental settings. The definition of a study design in-

cludes both a definition of the study base and of the study type. The main types of observational studies are the cohort study and the case-control study (sometimes called case-referent study). Although their theoretical background is the same, as in theory every case-control study takes place within a cohort (Breslow and Day 1987, p 3), the practical implications of conducting a cohort or a case-control study are quite different. A number of other designs have been used or proposed and are mostly variations of these basic types.

12.3.1 Study Base

Once the general objective of the study has been defined (at the conceptual level), and before the study type can be described, the investigator should identify the actual setting, or study base, in which the particular scientific problem can be studied. The study base should not only be a population (a number of individuals), but the morbidity experience of this population during a certain period of time (Hernberg 1992, Chap. 4).

The definition of the study base may depend on the study aims (testing a specific hypothesis or generating hypotheses). For example, the study base may be the 10-year follow-up of workers employed for at least one year in a particular industry and exposed to a specific chemical (hypothesis testing). Another example of a study base would be the one-year incidence of a disease among individuals living in a certain geographical area and having a wide range of exposures (hypothesis generating).

The study base should be defined in terms of eligibility ("employed at least one year" – "living in a particular area"), of its size, and of distribution of exposure, confounders and modifiers (as identified from the literature review). One should also define the time period during which information on morbidity has accrued ("10-year follow-up" – "one-year incidence data").

Ideally, a study base should make possible the scientific generalization of study results (the relation between exposure and disease). This can be achieved if exposure conditions, as well as potential confounders and effect modifiers can be measured within the study base. If not, the particular morbidity experience of the study population would be mostly descriptive and would apply only to the empirical population under study.

12.3.2 Cohort Studies

The design and analysis of cohort studies is described in detail in Chap. I.5 of this handbook. In cohort studies the study sample to be included is either the whole study base or a sample of it based on some existing exposure information. Thus unlike in a case-control framework, the complete assessment of the evolution of the morbidity in the course of the follow-up time is one of the challenges of this type of studies. This can be based either on a *passive follow-up* by matching the cohort with routine records of mortality or incidence coming from registries or

administrative sources (cf. Chap. I.4 of this handbook) or on an *active follow-up* by which the health status of every subject in the cohort is determined at several time points during the study period. On the contrary, at least in principle, the assessment of the exposure is less problematic than in case-control studies as it is determined before the disease arises. Conceptually, all cohort studies are prospective in the sense that one measures the exposure before disease occurrence. In practice this approach is often not realistic when the outcome of interest is a disease with a long induction period like cancer, so that a purely prospective observation of the morbidity would either imply a study duration of several decades or very large numbers of participants. One way to shorten this study time is to define the population historically, i.e. at a given time in the past and to mimic the prospective follow-up of the cohort until the present time. This design is called *a historical cohort* design in contrast to the *prospective cohort* design.

In cohort designs the main issue is the presence of study drop-outs or subjects *lost to follow-up*. When a drop-out occurs, its influence can only be controlled in the analysis assuming that the drop-out is only determined by recorded factors. In other terms, this means that we know for each missing subject, why she/he is missing and in particular that the drop-out does not depend on the unmeasured health outcome. In technically terms this is the MAR hypothesis (cf. Chap. II.6 of this handbook). This hypothesis can, however, not be tested from the data as – per definition of drop-out – we do not know the study outcome. An example of this would be a study in which the outcome of interest is smoking cessation comparing different cessation strategies. If the study participants fail to show up at later interviews aimed at investigating whether the subjects are still abstinent, this may be because they have relapsed and are ashamed of this fact or because smoking is no longer a problem as they have been abstinent for a long period. The effect of study drop-outs can then only be assessed through sensitivity studies by which one would simulate reasonable data for the missing study outcome. In the smoking cessation example one could for instance assume that all drop-outs relapsed, or a 90%, 70%, … random proportion of the drop-outs. Preventing study drop-outs is thus the main challenge in cohort studies.

Historical Cohorts

As mentioned earlier, the historical cohort design mimics a prospective follow-up of a historically defined population. This entails that at any point in the past, it is possible to check whether a subject satisfies the cohort-inclusion criteria as defined in the protocol or not. Furthermore if a subject from this cohort developed the outcome of interest, one must be sure to detect it and to be able to determine at what time point this event of interest occurred. In historical cohorts, the operational definition of loss to follow-up is the following: if the event of interest occurs at a given time-point for a given subject one must be sure to know it. If this cannot be assured from a certain point in time ownards, the subject must be considered as lost to follow-up.

The historical cohort design has been the design of choice for industry-based epidemiology of chronic diseases especially cancer (cf. Chap. III.2 of this handbook) and we shall discuss the relevant issues in this context.

Cohort Recruitment. In a historical cohort, the first step in planning is to determine whether information exists by which one can be assured that the population satisfying the theoretical cohort definition can be collected. Completeness with respect to the cohort definition is paramount. For instance if the cohort definition is "all subjects having worked on a given industrial site since 1970", a document dated from 1970 listing all those subjects must be available as well as yearly lists of all subsequently hired subjects. Computerized files can rarely be trusted as they were usually created for administrative purposes and are often at least partially overwritten or may exclude some categories of employees. Individual files are also to be dismissed as a single source of data. Neglect, lack of place for archives or floods and fire may have led to the selective destruction of files. It cannot be assumed that these lacking data are independent from either exposure or health outcome. In order to reliably identify a historical cohort, at least two data sources which can be considered independent should be available. One example would be separate lists from the pension scheme and from the personnel department. An alternative would be an enumerated list from whatever origin and historical documents in which yearly counts of employed subjects are given.

Case Ascertainment and Loss to Follow-up. In historical cohorts the main problem is to be sure to that any case that occurred in the follow-up period was detected. As an active follow-up is impossible retrospectively, the tracing procedure of subjects must rely on the matching of the cohort data with routine records. These routine records have usually a limited recruitment area. Disease or mortality registries are regional in most countries (with the notable exception of the Scandinavian countries) and sometimes related to the place of birth or residence. When either information is incomplete or missing for some subjects or if some subjects moved out of recruitment area, these subjects are lost to follow-up i.e. from that moment on, if the event of interest had taken place, it will no longer have been recorded with certainty. In the statistical analysis henceforth the subject should no longer contribute any person-time (cf. Chap. I.5 of this handbook). Therefore it is important to set up procedures by which one can determine whether and when a subject is lost to follow-up, for instance by trying to trace the addresses of the subjects through the pension schemes. Setting up such procedures can be difficult or even impossible for selected groups. For example historical cohorts may include foreign-born subjects who may have returned to their country of origin. If a subgroup is identified, for which such loss of follow-up is likely but no individual tracing can be set up, the only solution might be to drop all members of this subgroup from the study at the time of last contact. Thus in a historical occupational cohort study without active follow-up, one may need to consider as lost to follow-up all foreign-born employees from the date of last employment.

Recording individual diseases or medical causes of death may be restricted by data protection laws in absence of an informed consent. Strategies must then be set up, often involving third parties, by which the epidemiologists will have access to anonymous but individual data (e.g. Wild et al. 1995a).

Prospective Cohorts

As mentioned, the major advantage of prospective cohorts is that the exposure information is determined before the disease occurs, which allows to precisely measure the exposure and confounders of interest. In historical cohorts the exposure is estimated retrospectively although it relies as much as possible on historical exposure data. But these data were usually not collected for this purpose with the possible consequences with respect to their validity. This leads often to missing or imprecisely measured exposure data. In the worst case, such a retrospective exposure assessment may lead to an information bias not unlike that potentially occurring in case-control studies. In prospective cohorts, the information bias is theoretically impossible.

Another advantage is that as informed consent can be obtained for each study participant, there are no legal problems with respect to access to data.

Although in some very large studies as the American Cancer Prevention Studies (Garfinkel 1985), it is possible to study long latency chronic diseases, i.e. in general diseases with long induction periods, most prospective cohorts are primarily targeted at subclinical disorders assessed by questionnaires and functional or biological measurements.

An inherent drawback of prospective studies is their practical difficulty. Such prospective follow-ups of populations require repeated contacts with each subject of the cohort and are very cost-intensive.

A first problem is that the participation proportion is often rather low, especially if that participation entails repeated contacts which might discourage taking part. For instance, participation varied between 22% and 38% in the German centres of the European Prospective Investigation into Cancer and Nutrition (EPIC) study (Boeing et al. 1999). Limited information exists (Goldberg et al. 2001) which tends to show that participants are generally in better health than non-participants. Most prospective cohorts therefore rely on voluntaries. Thus prospective cohorts are only exceptionally representative of their target populations. Fortunately, representativeness is not a key issue in analytical studies as long as the loss to follow-up is limited. In prospective cohorts, loss to follow-up occurs either due to actual loss of contact or more often due to the refusal of continued participation. An unpublished survey by Moulin (personal communication) of all large ongoing prospective cohorts in France suggested that reasonable numbers of subjects lost to follow-up can only be obtained by regular contacts, regular feedback of the results of the studies to the participants and, if possible, presence in the media. When planning prospective studies, enough resources should therefore be allocated not only to the actual contacts with the subjects (i.e. mailing of questionnaires and reminders) but also

to the communication budget, both with regard to the media and the study participants.

12.3.3 Case-Control Studies

In a case-control study, incident cases of a given disease are gathered from the study base and are contrasted to a sample of controls drawn from the same base. Exposure histories are then collected from the cases and the controls. The theoretical foundations of the design of a case-control study (also called a case-referent approach as in Miettinen (1985)) are the same as for a cohort study (see also Chap. I.6 of this handbook). But instead of comparing the disease incidence between exposure groups, it compares the exposure between cases and controls.

According to the most frequent definitions, a case-control study conducted within a dynamic population may be *population-based* or *hospital-based* depending on the selection procedures of cases and controls. When the base population is an enumerated cohort, the case-control study is often called *nested within the cohort*. A case-control study designed for etiological research will belong to one of these three categories. Each of them has different practical implications which we detail below.

Population-based Case-Control Study (Primary Base)

In a population-based case-control study, the cases are all patients diagnosed with the disease during the study period among those who live within a country or a region. The controls are sampled from this population. In this design, the study base, i.e. the population living in this country or region during the study period, is precisely defined a priori before the start of the study. Another example of a primary-base case-control study would be to include all members of a given health care insurance system where all cases are also members.

Case Recruitment. A researcher may choose a population-based case-control design if all new cases of the disease (or at least a representative sample of all cases) can be identified in the study base. As it is essential to ensure completeness of case-finding, a disease registry may be helpful for identifying the cases. The existence of a disease registry may also constitute a motivation for conducting a study in a given area. However, the value of a registry may be limited if there is a substantial time lag between diagnosis and registration, particularly if collection of data from the respondent is necessary and if the disease is rapidly fatal. Thus, a specific procedure for case identification has to be set up in most studies by the research team. For example, a case-finding network may be organized including all hospitals, clinics, and pathology departments in the study area to identify and interview the cases. This should also be extended to nearby areas as some of the diseased persons in the source population defined by residence in a given geographic area may be diagnosed and hospitalized elsewhere. It is also strongly recommended to perform an active search of the patients, by organizing periodic

visits to all centres, rather than to rely on a passive notification of the cases by the medical staff of the clinics or the hospitals.

Selection of Controls. *Principles.* The controls should be selected at random from the same base population as the cases. In addition, the probability of selecting a particular control subject should be proportional to the amount of time that he contributes to the study period, or to person-time at risk (Rothman and Greenland 1998). For example, if a subject moved out of the source population at half the study period, he/she should have only half the probability of being selected as a control than a subject who stayed in the source population during the entire study period. To be eligible, a control subject should belong to the study base at the date of diagnosis of the index case. Controls who recently moved into the source population and are chosen to match cases diagnosed several years earlier should be excluded since they are outside the study base. Excluding controls who have recently moved in the base reduces the problem, but does not solve it, since people who have moved out of the base will still be missed.

A density sampling of the base population should be used. Density sampling can be achieved, for example, by selecting the controls at a steady rate throughout the study period proportional to the number of cases. In practice, the protocol may define several points in time, e.g., once a month or once a year, where controls will be selected from the population present in the study base at that time. In a design where the cases and the controls are matched individually, it is also possible to use sampling sets of possible controls, one per case, composed of all persons present in the source population at the time of case's diagnosis. The desired number of matched controls is then selected at random within each of these risk sets.

Selection of Population Controls Based on a Listing of Individuals. To be feasible, these procedures of control selection necessitate not only a fully enumerated source population, but also regular updates of this population, to take emigration and immigration during the study period into account. In Scandinavian countries, study investigators may rely on central population registers to select controls using a simple random sampling at regular intervals during the study period (see also Chap. I.4 of this handbook). More often, however, a complete population register, including the identification of individual members by name and address, as well as stratification variables such as gender and date of birth, will not be available and other methods of control selection must be used.

In the absence of a population register, the researcher may use other *lists of individuals*, such as lists of municipality residents, electoral lists, telephone books, listings of health insurance members, and so forth. Using these lists for control selection, however, may introduce bias if the probability for an individual in the source population to be listed is related to the exposure of interest. Telephone books would not be suitable for a study on cancer in relation to an occupational chemical, for example if phone numbers of highly educated and less exposed individuals are less frequently published in the directory than phone numbers of subjects of other socio-economic categories. Persons not registered in electoral

rosters may also differ from those listed, and may not include immigrant workers from foreign countries. Municipal lists may not be updated regularly. The decision to use such lists for control selection must be done carefully. Beside completeness of the list, the possibility of tracing individuals based on the information provided (name, address, phone number, etc.) must also be considered. One should also check that the cases are listed on the roster. The analysis should then exclude the unlisted cases, such as those who are not citizens when using electoral lists.

Selection of Population Controls Without a Listing of Individuals. Other sampling schemes may be used for selecting the controls when no list of individuals is available. Multistage random samplings starting with sampling of dwellings or based on random digit dialing procedures are commonly used. The controls can then be selected within each household.

Neighbourhood controls. This method implies a two-stage sampling, with a random sampling of households followed by the selection of an eligible individual within the selected residence. Households sampling may be conducted from a roster of residences, obtained for example from census data. When a roster of households is not available, controls may still be selected from residences in the case's neighbourhood. Starting from the case's residence, the interviewer may follow a predefined procedure for selecting a household, by means of a map, aerial pictures, or by a systematic walk algorithm starting at the index household (Wacholder et al. 1992b). This sampling method implies that the controls are individually matched to cases on place of residence. To avoid bias, the interviewer should not be given the flexibility to choose which house to select but a simple and unambiguous algorithm for selecting households should be developed to remove the possibility of interviewers avoiding certain areas. A potential problem of neighbourhood controls is overmatching on the exposure, due to similarities between cases and controls living in the same neighbourhood.

Random digit dialling. Random digit dialling can be used to select population controls when no roster exists and when almost every household has a telephone. Random digit dialling generates telephone numbers, and it does not rely on a telephone book where new or unpublished phone numbers are not listed. Several variants of the standard method exist (Waksberg 1978). Briefly, a phone number is created using the first numbers, including area code, of working telephone numbers provided by the telephone company, which are then completed with random numbers. The number is dialled a predetermined number of times. The first contact with a household member is used for screening and to obtain a census of the household. Based on the responses, and a predetermined sampling scheme, eligible subjects are selected to be controls. These individuals can be interviewed by telephone directly, or they can be contacted afterwards for an in-person or a telephone interview.

Random digit dialling is not appropriate if the telephone coverage is low, but this should not be a problem in most developed countries. Other problems associated with random digit dialling include residences that can be reached by

more than one phone number, or if more than one person in the household is eligible to be a control, since it may lead to different selection probability of the controls. It should also be realized that random digit dialling is an expensive and time-consuming procedure, particularly when targeting subgroups of the population, since a large number of phone calls may be necessary before the desired number of eligible controls are found. Non-response and refusal is an additional problem and it may not be possible to have an exact estimate of the participation rate. It is recommended that the distribution of the final sample according to key variables such as age, sex or socio-economic status is compared to an expected distribution obtained for example from the last census. The random digit dialling has been used successfully in a large number of studies, but new technology, such as the widespread and sometimes exclusive use of mobile phones, may cause this method to be obsolete in the near future.

Hospital-based Case-Control (Secondary Base)

In hospital-based case-control studies, the cases are the patients diagnosed with the disease in a given hospital or clinic during the study period. The controls have to be selected from the population from which the cases arose, i.e. the group of individuals who would be treated in this hospital if they had developed the disease. Because the source population is not easily identified, a random sample of controls can hardly be selected directly from this population. Instead, it is usually more practical to select controls among patients with other diseases diagnosed in the same hospital, representing a non-random subset of the study base. An appropriate group of control patients should have the same referral patterns to that hospital than the cases, so that the controls would have been admitted to this hospital if they had the case disease. The possibility of selecting hospital controls rests on the assumption that they are representative of the exposure distribution in the source population. This assumption is reasonable if the control disease is not causally related to exposure, and if exposure is not related to admission to that hospital.

Case-Control Study Nested Within a Cohort

When the study base is a real enumerated cohort with available entry and exit times a case-control study may be conducted by drawing cases and controls from this cohort as the source population. It is called a "case-control study nested within a cohort". Using this design implies that the cohort has been constituted so that the controls can be selected adequately from the cohort. As this design relies on an enumerated source population, a control group can easily be identified. In a matched design for example, a set of possible controls can be constituted with all non-diseased individuals in the cohort at the time of the cases' diagnosis. One or several controls may then be selected at random from each set.

12.3.4 The Study Base Principle: Selection, Exclusion and Resulting Bias

The study base principle, the first of three principles Wacholder et al. (1992a) developed for the selection of controls in case-control studies, that also applies to the design of cohort studies, simply states that "cases and controls should be representative of the same base experience." (see also Chap. I.6 of this handbook). *Representativeness* for the general population is not needed in analytical studies of the relation between an exposure and a disease. In a representative population, an association that is limited to one group may be obscured because the effect is weaker in other groups.

Thus for the aim of scientific inference of the relation between an exposure and disease, any exclusion or inclusion criteria are valid as long as they apply equally to cases and controls.

Wacholder et al. (1992a) identified the following reasons for which exclusion criteria can be applied:

- Inconvenience: Subjects of a given subgroup might be hard to reach. Failure to exclude a priori such a group may lead to a very poor response rate and an a posteriori exclusion, with the corresponding waste of resources.
- Anticipated low or inaccurate responses, e.g. of subjects who do not speak the language of the interview sufficiently well. Failure to exclude a priori such a group may yield non-interpretable data.
- Lack of variability of the exposure: If one intends to set up a cohort investigating the dose-response effect of a potential occupational carcinogen like cobalt salts, inclusion of a large number of workers from industries which do not use these chemicals does make little sense although including a small group for stabilising the baseline category may still be justified.
- Subjects at increased risk of disease due to other causes: In a prospective study targeted on environmental effects on asthma, subjects at high asthma risk due to their occupational exposure (e.g. bakers) should be excluded because cases are likely to be attributable to the occupational exposure and therefore may not contribute to the understanding of other risk factors.
- Combination of the above: In a historical occupational cohort study based on a factory, short-time employees may be difficult to track, their exposure is likely to be determined much more by previous or subsequent work, their cumulative exposure within the company is bound to be low and they may be at increased risk for many diseases as they constitute a group of socially unstable workers who are likely to have other risk factors. It is thus standard in such settings to exclude short-time workers. This has, however, the consequence that any given subject of the study base contributes person-time only from the date on when he/she has reached the minimum employment duration.

On the other hand, if a study base is restricted the comparison with the general population is biased. The so-called Healthy Worker Effect, by which is meant that a series of extraneous factors usually lower the observed mortality among

employed workers, is an example of a selection bias. A simple comparison of the mortality of a cohort with population mortality rates as expressed for instance through the standardized mortality ratio is thus of limited validity. Of course, as discussed above, internal comparisons of exposure groups are still valid but may lack the necessary power for useful scientific inference except in very large cohorts.

Another consequence of severely restricting the study base of a (cohort or case-control) study is that it can lead to reduced detection of variability of the strength of association (effect modification). If the effect of smoking were to amplify the effect of an environmental exposure, restricting the study base to non-smokers may lead to a spurious absence of effects.

In some settings representativeness is an issue. If the study is focused on attributable or absolute risks, the study must be either exhaustive or representative, or external information must be available with respect to the sampling fractions of the strata of the study population within the general population.

Choosing Between Epidemiological Designs 12.3.5

Many other etiological designs like case-cohort designs, case-only designs, two-phase sampling, counter-matched designs to cite just a few, have been proposed in the epidemiological literature (Wacholder et al. 1992c; Chap. I.7 of this handbook). These designs usually combine elements of the case-control and the cohort designs often either by making additional hypotheses (case-only design) or by making use of additional data like in two-phase designs (see e.g. Breslow and Chatterjee 1999).

Another design is the so-called ecologic design, in which the units are groups of people rather than individual subjects. The groups may be classes in a school, factories, cities or administrative areas within a country. The only requirement is that a measure of the exposure and disease distributions is available for in each group. Because the data in ecologic studies are measurements averaged over individuals, the degree of association does not reflect the association between exposure and disease among individuals (so-called ecologic bias, see Greenland and Robins (1994)). Thus, while ecologic studies can be useful for detecting associations of exposure distributions with disease occurrence, such a design should not be used for etiologic investigation.

Another possible design consists in selecting a cross-section of the study base with no time dimension (cross-sectional study). Here both exposure and disease status are collected simultaneously at one point in time. It is therefore not the incidence but rather the prevalence of the disease that is investigated, so that it is usually impossible to assess whether the exposure actually preceded its presumed effect on health. Moreover, cross-sectional studies are particularly prone to selection bias as diseased subjects may have left the study-population. A longitudinal observation of exposure and disease in a study that is the paradigm of both the cohort and the case-control design is better suited for solving etiologic problems than a cross-section of it (a prevalence study). Possible exceptions are diseases with short induction periods or exposures that cannot change such as blood type or other invariable personal characteristics (see Rothman and Greenland 1998,

p 75). If, however, a prospective cohort design is not feasible for financial reasons or simply because it is impossible to follow up a large enough group of subjects, the cross-sectional design can, despite its above mentioned limitations, provide important information especially if targeted on several (non-fatal) health outcomes (see Wild et al. 1995b) or if a number of different exposures coexist.

The main study types remain the cohort and the case-control designs. Choosing one or the other design depends on a number of issues among which the incidence rate of the disease of interest and the prevalence of the exposure of interest are prominent.

If the disease is very rare (for instance a rare cancer as testicular or brain cancer), a cohort approach would necessitate huge numbers of participants and a long follow-up to identify enough cases to make useful inference. A case-control study is the more reasonable approach for rare diseases. Another situation where the case-control is the preferred approach is if the aim is to generate hypotheses concerning various exposures in relation to a given disease. Determining possible occupational origins of laryngeal cancer is such an example.

On the other hand, if the exposure is rare and restricted to easily identified subpopulations or if one is interested in all possible disease outcomes of a given exposure, a cohort study, either historical or prospective, is the most efficient choice. An example for the former reasons would be a study of the carcinogenic effects of hard metal dusts (Moulin et al. 1998b) which occurs mostly in the factories producing hard metal tools. An example of the latter would be if one were to study the health consequence of the occupational stress in call-centres.

Another aspect that can influence the choice of a design is the induction time of the disease (although it may depend on the exposure). For long induction periods, a prospective cohort study is clearly not the design of choice as this would imply waiting a long time for a sufficient number of events to occur. This problem can be circumvented by the historical cohort design. Most historical cohorts are focused on cancer, a long induction disease per se. The choice between a case-control and a historical cohort study is then dependent on whether (or not) a historical cohort can provide an answer to the research question.

Other issues which might influence the choice of a given design include the precise scientific aim of the study. Direct estimation of the population incidence would require enrolment of the target population that ideally needs a cohort design or at least a population-based case-control approach. If, on the other hand, one is interested in the precise temporal sequence, as for instance in the study of the evolution of CD4+ cell numbers in HIV infected patients (Kaslow et al. 1987), cohort studies are virtually the only available design which may, however, be supplemented by a nested case-control study.

The final choice, once a theoretically optimal design has been determined, depends on the actual feasibility of each study as well as on the practical terms of access to the data and the costs involved.

Table 12.1 summarizes strengths and weaknesses of the above study types.

Table 12.1. Strengths and limitations of different observational study designs

Investigation of …	Study design		
	Cross-sectional	Case-control	Cohort
Rare diseases	–	+++++	–
Rare causes	–	–	+++++
Multiple endpoints	++	–	+++++
Multiple exposures including confounders	+++	++++	(+++)[a]
Temporal sequence of exposure and disease	–	(+)[b]	+++++
Direct measurement of incidence	–	(+)[c]	+++++
Long induction periods	+	++++	(+++)[d]

Suitability of study design: +++++ highly suitable; ++++ very suitable; +++ suitable; ++ moderately suitable; + limited suitability; – not suitable

[a] If prospective (multi-purpose cohorts)

[b] If nested in a cohort

[c] If population-based, combined with an incidence study

[d] If historical

Statistical Power

12.3.6

As mentioned in the first section any study protocol should include an evaluation of the statistical power to detect a predefined effect of the exposure with a given study base. When computing the statistical power, the need for subgrouping, based either on exposure classes or confounder classes, is important. In practice the choice of the sample size (or to be exact the size of the study base) is a compromise between what one would ideally be able to detect and the practical limits of the study. These limits are of two kinds. A first limitation occurs when there simply do not exist enough subjects in the envisaged study with either a rare disease of interest e.g. rare cancers, or with a rare exposure. An example of this would be if we were to study the interaction between a rare gene (prevalence < 0.01) and a rare environmental exposure. If the latter has a prevalence of 5%, less than 5/10,000 of controls would show both features so that the minimum sample size to investigate such an interaction would be in the tens of thousands. A second limitation is when the needed funding is not reasonable.

In general, a statistical power of 80% is considered a reasonable power to achieve objectives and any power below this arbitrary figure may be considered too low. If methods of prior assessment, either formal or intuitive, suggest that the study will be too small to be informative, there are several options:

One can lower the level of ambition. For instance instead of computing a statistical power to detect a 1.2 odds ratio, this value can be put at a 1.3 level. The drawback of this strategy is of course that if no effect of the exposure can be detected, lower risks cannot be excluded. Thus if the main interest lies in effects of low doses of exposures with expected low magnitude effects such a strategy is not recommended.

One can set-up a system with an extended follow-up time. This means that the study base is enlarged in its time dimension. This usually entails that the results will be available later.

One can try to organize a multi-centre study. This means that the study base is enlarged in its number of subject dimension. When organizing multi-centre studies, one should be reasonably sure that the gain in sample size is not offset by between-centre differences in exposure circumstances and assessment, or differences in case ascertainment. Another issue in this case is that the harmonization of centres has its costs, too.

One can abandon the project. It can be considered unethical to undertake a study which would not add to the general knowledge but costs money which could benefit other research.

Statistical power can, however, not be the only criteria by which to judge the appropriateness of a study. Other reasons like public concern or scientific background knowledge for instance based on positive animal studies are sometimes even more important.

12.4 Measures of Disease Outcome and Exposure Parameters

12.4.1 Measurement and Classification of Exposure

Introduction

Epidemiological studies are designed to assess the impact of exposure on the development of a disease. The sources of error and the ways in which exposure and disease are assessed are quite different, and thus the mechanisms by which errors arise are different as well.

The range of exposures of interest in epidemiology is broad (Savitz 2003). Exposure include exogenous agents such as drugs, diet, and chemical or physical hazards present in the environment; genetic attributes that affect ability to metabolize specific compounds; stable characteristics such as height or eye colour; physiologic attributes such as blood pressure; life habits such as physical exercise or tobacco smoking; mental states such as depression; social environment, and so forth. To this wide range of exposures of interest correspond many different methods for measuring exposure (Armstrong et al. 1994; Chap. I.11 of this handbook).

The method chosen to collect data depends on the particular exposure to study, the precision of data required, availability of existing records, sensitivity of subject to questioning about the exposure, cost of various methods, etc. The study protocol should describe the operational approach chosen for exposure ascertainment. The accuracy of an operational approach is best described in relation to the ideal measure which is always impractical or impossible to obtain.

The ultimate goal of exposure assessment is to measure the exposure that contributes to the etiologic process under investigation. Ideally, an exposure assessment should take this biologically effective exposure, and only this exposure, into consideration. Most often, however, this goal cannot be reached by an operational exposure indicator. One reason is that the biological mechanism by which exposure might cause disease is often not clearly identified. For example, in studies on the potential cancer risks associated with exposure to ELF magnetic fields, it was not clear whether the most relevant exposure indicator with respect to aetiology should be an average exposure over the entire life or over the most recent period before cancer diagnosis, or if it should be a measure of peak exposures over a certain threshold, or in a particular exposure windows. This problem could be partially overcome if a detailed exposure profile over time can be estimated, to calculate different exposure indices, each of them being related to cancer risk. Another reason is that it is often impossible to obtain the data that would reflect perfectly the biologically effective exposure. For example, the persistent organochlorine pesticide DDT and its metabolite DDE were suspected to be causally related to breast cancer risk (Wolff et al. 1993). Assume also that the biologically effective exposure is the level of DDT/DDE present in breast tissue in the time window 5–15 years before cancer diagnosis. Different exposure measures among cases and controls could be used to study the relationship between DDT/DDE exposure and breast cancer (Savitz 2003), including environmental levels measured in the area of residence at the time of diagnosis or taking into account the residential changes of the subject, present-day blood levels, or blood levels measured in the etiologically relevant period using serum specimens drawn in the past and kept in a bank. Clearly, these exposure measures are not equivalent as they correlate differently with the biologically effective exposure. Using environmental exposures as a surrogate exposure indicator is probably ineffective, since the association on breast cancer risk may fall below what can be detected. However, if blood levels are strongly correlated with breast tissue concentrations, an association with breast cancer risk can still be observed, although blood levels are not the right measure. Nevertheless, certain metabolites measured in blood may be a good indicator of past exposure if the half life of the agent is sufficiently long (Flesch-Janys et al. 1998). The choice of an operational indicator of exposure should be done in the context of a hypothesis on biological process, and a comparison of the operational and ideal exposure indicators should be provided.

All laboratory data are subject to error due to imprecise measurement. However, the conceptual error by which the measure obtained does not reflect the exposure of interest is often of much greater importance. In a study focused on effects of microbiological contamination, measurement of viable colonies may be only of marginal interest if these bacteria are not the pathogenic ones. Bacterial endotoxin has been shown to be the more relevant exposure with respect to lung diseases (Rylander 2002).

Another challenge in exposure assessment is that often different types of exposure coexist and their individual effect can only be considered in combination. The level of exposure aggregation that is chosen must, however, reflect scientific hy-

potheses. For instance a nutritional epidemiology study focused on the role of coffee in miscarriages could consider two relevant categorizations. The role of caffeine itself would be investigated by grouping all sources including tea, caffeine-containing medications, etc., whereas the role of constituents of the coffee other than caffeine would be investigated by grouping caffeinated and decaffeinated coffee.

Temporal Aspects

Some exposures are constant over time, such as genetic constitution, but all exogenous exposures such as diet and chemical pollutants vary substantially over time. In addition to the identification of a biologically relevant exposure, it is necessary to identify an etiologically relevant time window during which the exposure may be related to disease occurrence so that the data collection concentrates on etiologically relevant time-windows.

It has been long recognized that many diseases such as cancer appear a long time after they have been induced. The time between the beginning of the exposure and the first manifestation of the disease of interest is called the *induction* period. It serves as a surrogate for the biological induction period in epidemiological studies although the onset of exposure does not necessarily result in immediate induction. Epidemiological studies should allow for the fact that diseases with long induction periods that appear immediately after the exposure are not attributable to this exposure. In the statistical analysis of such data, the usual practice is to shift the exposure by a given *lag time* which is typically about half the usual induction period. This has the consequence to ignore the exposure of recent years. The consequences for planning is that whenever the anticipated induction period is long, the investigator must select a study base including a sufficient number of subjects for whom the exposure is ancient enough. On the other hand, if the expected effects of exposure are of a short-term nature as for instance a reversible genotoxic effect as assessed by a comet assay, it is important to precisely measure the relevant short-term exposure.

Other time-related issues may concern the temporal pattern of the exposure itself and its presupposed effects. If the exposure effect occurs through the action of peak exposure, its effects are likely to be much more important when the exposure is highly variable than in circumstances in which the exposure is virtually constant. For such presupposed effects, the exposure measurement of most interest may be an estimated number of peak exposures. On the other hand if we assume that the exposure acts through a cumulative damage, the total cumulative exposure is the adequate way to express its effects. If the disease of interest is reversible, the cumulative exposure may also be inadequate as a measure of exposure and exposure assessment in different time windows may be more relevant.

Sources of Exposure Data

The different possible sources of exposure data and their characteristics are described in Armstrong et al. (1994). The following section is mostly a summary of the issues covered in this book (see also Chap. I.11 of this handbook).

Questionnaires. A prominent and often the only way to assess an individual exposure is by an exposure questionnaire. The three main types of exposure questionnaires are self-administered questionnaires, telephone administered interviews and interviewer-administered questionnaires, also called structured personal interviews. The last two methods are the commonest methods of collecting data on exposure in epidemiological studies. Using them allows the number of errors and missing items to be reduced and more complex information to be obtained. For instance, there is a possibility of branching, i.e. specific questionnaires can be inserted after some trigger information has been recorded. This possibility has been used in occupational exposure questionnaires in which job histories are obtained and specific questionnaires are used for a limited number of jobs and/or tasks (Ahrens et al. 1993). On the other hand, the interviewer may increase error if he or she exerts a qualitative influence on the subject's responses by his or her appearance, manner, method of administration, etc. Intensive training and standardization of the interviewers is therefore very important in the planning and conduct of a study. If a series of prerequisites are fulfilled as the stressing of interview neutrality in training, the standardization of questionnaire wording and its administration, the likelihood that the interviewer's personal attitudes will affect the responses is much reduced. The main advantage of a self-administered questionnaire is its reduced cost. Armstrong et al. (1994, p 44) conclude that "there appears to be little difference between these methods (subjective recall of exposure collected through face-to-face, telephone or self-administered questionnaires) with respect to the validity of the data obtained. (…) Face-to-face interviews are the dominant approach and are clearly best for the collection of large amounts of complex data. However, where subjects are widely dispersed and the questionnaire can be kept comparatively brief, telephone interviews can be favoured. Self-administered questionnaires should be considered for low budget studies for which small amounts of reasonably simple data are required."

In setting up a questionnaire many sometimes contradictory issues arise. While obviously more details can theoretically be assessed if it is longer, long questionnaires take more time and (especially among diseased subjects) may be more difficult to apply. A questionnaire should have a clear structure and should be understandable. Questions like: "Have you been exposed to bischloromethylether?" should be avoided. Sensitive issues (religion, sexual habits, alcohol consumption, etc.) should be avoided if they are not central to the study and should be given careful consideration if necessary in order to avoid withdrawal of the interviewee. A questionnaire should always be tested before use. The use of validated instruments is of course desirable. Correspondingly, a large number of publications exists on validating existing questionnaires (see for instance Rouch et al. (2003) or Bogers et al. (2004)). It is probably a good strategy to choose, whenever possible, an already existing questionnaire that has been validated in circumstances close to its intended use. If one decides, nevertheless, to adapt a questionnaire for a given study, a pre-test should be carried out to investigate its properties, notably feasibility, clarity, and reproducibility, so that necessary adaptations can be made before applying it. If a completely new instrument has to be designed a validation study should be taken

into consideration to investigate its properties as e.g. validity and responsiveness to change, i.e. its ability to reflect changes in behaviour or subjective symptoms (Bogers et al. 2004).

Diaries. Diaries refer to detailed prospective records of exposure by the subject. As such they can neither be used in case-control studies nor in historical cohort studies. This method has been used in many contexts among which their use in nutritional epidemiology for measuring dietary intake is prominent (cf. Andersen et al. 2004; Chap. III.4 of this handbook). Ongoing monitoring of symptoms of a disease is another application of diaries (cf. Goebel et al. 2002). Armstrong et al. (1994, p 219) conclude "The use of diaries may be highly accurate method of measuring present common behaviours. The limitations of diaries, in comparison with interview methods, are the greater burden on subjects, which may lead to poorer response rate and the greater cost for subject training and for coding the data. The accuracy of diary information can be enhanced by use of multiple diary days spread over a sufficient time period, and by careful training of subjects and coders."

Records of Exposures. Historical records of pre-existing data may be a valuable and sometimes the single source of early exposure data. Two types of records can be useful for exposure assessment. A first type of records contains information on the individual study subjects, for instance medical, behavioural (smoking), or physical characteristics (weight) contained in medical records but also social or occupational data contained in population registries. A second type of records contains information relevant to the exposure of groups like descriptions or measurements of environmental exposure or descriptions of histories of industrial processes in occupational epidemiology. The primary advantage of records is that they can provide prospectively recorded information, collected on information in the past. For example, use of pharmacy records in a case-control study of prescription drug use could overcome lack of recall. Exposure assessment based on historical records is immune to information or recall bias. Most important in this context is the use of data from earlier cross-sectional surveys. The main drawback is the lack of control over the availability of records for each subject and their standardized recording and the difficulty to assess the validity of the existing data.

Biological Measurements and Measurements in the Environment. In principle, measurements made directly in the human body represent the ideal approach to measuring exposure for etiological studies. In practice a number of problems exists. A first problem is to determine the measurement in terms of the correct metabolite and the appropriate point in time that is relevant for the presumed biologically effective dose. A second problem is the often large within-person variability which can be of the same size or larger than between-person variation. This has been observed repeatedly in industrial exposures (see Kromhout et al. 1994), but it is often also true for biological measurements. Liu et al. (1978) reported a ratio of within-person to between-person variances as high as 3.20 for 24-hour urinary sodium, a marker of sodium intake.

The current fast development of methods of measurement of exposures in biological materials may, however, give rise to many useful indicators. The epidemiologist should be aware of the developments in this area. Methods whose validity have been assessed should always be preferred and the operations manual should include Good Laboratory Practices. In planning laboratory work, it may be important to keep track of the internal quality control procedures and provision should be made that these informations are recorded.

Objectivity of measurements in the environment are best achieved by personal sampling over extended periods of time. Relying on samples collected over relatively short periods of time without clear sampling design may induce substantial error in the exposure assessment. While sensitivity of measurements can be very high, the measurement error is often only a small part of exposure variance which is usually dominated by the intrinsic variability in exposure. Planning exposure measurements for an epidemiological study should therefore rely on factors likely to influence the exposure (Sauleau et al. 2003). The methods for exposure measurement have to be included in the operations manual. Reliable past exposure measurements rarely exist in sufficient numbers. Moreover, their validity is in general doubtful as they were usually obtained for purposes other than an epidemiological study, typically environment control. In such cases an attempt may be made to estimate exposure by use of conversion tables such as job-exposures matrices and food tables linked to data derived from records, questionnaires or expert assessment. Moulin et al. (1998b) show an example where existing past measurements were highly variable and only scarcely related to the exposure assessed by experts. Nevertheless, using the latter, they were able to demonstrate a dose-dependent carcinogenic effect of hard-metal dusts.

Finally the use one intends to make of the exposure measurements in the analysis of the data has to be stated in the protocol. This is required in order to assess whether a certain number of measurements with a given precision and a presumed variability of exposure will be sufficient to achieve statistically valid results with respect to the association of exposure and outcome.

Measurement and Classification of Health Outcomes 12.4.2

The main distinction to be made concerning measurement and classification of health outcomes is between diseases for which a clear diagnosis can be made at a precisely defined time and health outcomes which are defined by a measurement either by questionnaire or a functional or laboratory test.

The latter case includes health outcomes like obesity or hypertension where the health outcome is better expressed on a continuous scale rather than as a classification into diseased/non diseased categories and for which no clear-cut incidence date can be obtained. This feature implies that the prospective cohort design is the only design in which it is possible to be sure that the exposure precedes the disease. In such cases determination of the point in time when to measure the health outcome is crucial. If one is interested in acute effects of an exposure, e.g. the immediate effect of the chlorine in a swimming pool on genotoxicity using the

Comet assay, the health effect must be measured immediately after the exposure. If one is interested in chronic effects of an exposure, e.g. of organic solvents on chronic neurotoxicity, care must be taken to measure the health outcomes after a period of washout as for instance after the weekend in the example of a study of chronic neurotoxic effects of solvents, as the acute effects of the exposure may confuse the chronic effects. Such considerations may have serious implications on the planning, cost or even feasibility of such studies.

In the case of a well-defined disease with a defined incidence date, both historical cohorts and case-control designs can be used. Still the precise definition of the disease is one of the challenges of a clear design. The main issue is whether to group or to separate disease subtypes. Different subtypes of a same disease may have different risk factors even within a cancer site. The recent rise in adenocarcinoma of the lung has for instance been related to the increased use of "light" cigarettes. However, a too narrow definition of a disease may lead to smaller number of cases. This is even more an issue when the diagnosis is obtained from registry data or death certificates as is usually the case in historical cohorts. The precision of such data may well be fictitious and grouping of diseases thought to have similar etiologies may be the only reasonable choice.

An intermediate case would be a well-defined disease like COPD (chronic obstructive pulmonary disease) for which no systematic recording of patients is possible. For such a disease only hospital-based case-control studies with the challenge of estimating the time of onset of the disease in order to ignore all posterior irrelevant exposure, or prospective cohorts with the problem of potentially low power seem to be realistic designs.

12.4.3 Measurement and Classification of Health Outcomes

12.4.4 Information Bias

A major bias related to data collection is the information bias that occurs if the exposure assessment is different for cases and controls or the health outcome is measured differently for exposed and non-exposed subjects. The first situation arises mainly in case-control studies when the cases know of the possible determinants of their disease (e.g. smoking or asbestos exposure among lung cancer cases) and therefore report more of their past exposure. It is possible to minimize this type of bias by standardized questionnaires and, if possible, by a data collection blinded with respect to the case-control status. The latter approach, although an option in hospital-based case-control studies, is, however, virtually impossible in population-based case-control studies. The same type of information bias is possible with historical cohorts if information is obtained a posteriori from proxies.

An information bias can also be due to the unavoidable measurement error in retrospectively measuring the exposure. The comparable accuracy principle (Wacholder et al. 1992a) states that the degree of accuracy in measuring the exposure of interest for the cases should be equivalent to the degree of accuracy for the

controls unless the effect of the inaccuracy can be controlled in the analysis, as for instance by using appropriate validation data. Although adherence to this principle does not eliminate the corresponding misclassification, its rationale is to avoid that a positive finding is induced simply by differences in the accuracy of information about cases and controls (see also Chap. I.6 of this handbook).

12.5 Confounding

The *deconfounding principle*, another principle given in Wacholder et al. (1992a), states that confounding should not be allowed to distort the estimation of an effect. Confounders are per definition factors that are determinants of the disease and that are related to the exposure of interest. Setting up a list of confounders is therefore a primary task in the exploration of the scientific knowledge related to the research question of the planned study. This identification of confounders cannot only be based on statistical associations but must be thought in terms of potential causal pathways. For instance, if one were to study the effect of nutrition on coronary heart diseases (CHD), a factor that might be considered a potential confounder is obesity as it is related both to nutrition and CHD. On the other hand one might consider that obesity is on the causal pathway between nutrition and CHD in which case controlling for obesity would bias the estimation of the effect of nutrition (*overadjustment*). Careful thought should therefore be given to each factor that has to be included as a confounder. An operational result of this first step would be a list of three groups of confounders.

1. Established confounders that are known to determine disease and to be related to the exposure of interest but which are not on the causal pathways. Sex and age are virtually always included in this group.
2. Probable confounders that are established risk factors and for which there are reasons to believe that they might be related to the exposure.
3. Possible confounders including other risk factors of the disease that might be or not related to the exposure. If a given risk factor is a strong determinant of the disease (e.g. smoking for lung cancer), it might confound the association of a potential risk factor although the confounder is only weakly associated with the other risk factor in the study sample.

At the design stage, controlling for confounders can be done either by restricting the data base to certain values of the confounder (e.g. a lung cancer case-control study among non-smokers) or by matching. The latter ranges from broad frequency matching on age and sex to the individual matching on factors thought to be related to non-measurable factors (e.g. neighbourhood matching). It must be stressed at this point that controlling for confounders in the design may well be counterproductive as it is irreversible and may forbid to explore interesting post hoc hypotheses (see also Sect. 12.3 on matching in Chap. I.6 of this handbook). A last strategy is to collect the relevant data on confounders in order to be able to control them later in the analysis.

Details of how to deal with confounding using any of these methods can only be decided within the context of a given study. The main pitfall in controlling for confounders at the design stage is overmatching. Restricting the variability of the confounding variable will also reduce the variability of the exposure of interest within each matched case-control pair and will thus reduce the power of the study. This would occur if cases and controls were matched on age, sex and tight socio-economic categories. It is very likely that other variables like smoking or dietary factors would also be more similar.

12.6 Statistical Analysis

The statistical analysis of a study is an important step in the overall quality of a study and enough time and human resources should be planned from the start. Many large-scale studies we know of, are underreported because of lack of funding for the statistical analysis. It is impossible to detail all statistical analyses to be done at the planning stage. However, as already mentioned, the main research questions should be formulated in the protocol and these questions should be operationalized already at this stage, i.e. translated in a statistical hypothesis to be tested.

It is helpful at this stage to draw hypothetical causal graphs in order to formulate a priori models to be confronted with the data of the study (see Pearl 2000, Cox and Wermuth 1996). Figure 12.1 shows a very simple graph formulating a series of hypotheses on the effect of shiftwork and age on cognitive performances which may be mediated by sleep problems.

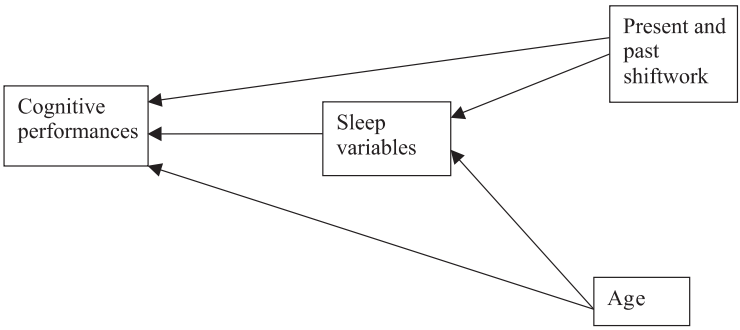

Figure 12.1. Causal hypotheses of effects of shift work and age on cognitive performances

Such a conceptual framework helps to set up a list of statistical analyses based on a priori hypotheses and additional ad hoc analyses which are more or less data driven. The minimum statistical content to be defined at the planning stage are the endpoint variable(s), the exposure variables of interest and the variables to be adjusted for (i.e. the potential confounders). The statistical measures (proportions, odds, means, standard deviations, etc.) and models (mostly regression

models) to be calculated should also be enumerated including the residual and influence statistics to be applied to check the robustness of the results. Finally, all the subgroups of intrinsic interest for which separate analyses will be done should be identified.

The fact that the a priori formulated hypotheses are operationalized, means that for instance the main endpoint(s) of interest should be identified as certain values of given variables. This is relatively straightforward in case-control studies for which the endpoint is the case status whatever its definition. For cohort studies focused on the mortality or incidence of given diseases, the list of diseases possibly related to the exposure of interest should be identified. For prospective cohorts or cross-sectional studies the endpoint of interest may be less straightforward and can possibly imply a prior data transformation, e.g. a score of depression obtained from a mental health questionnaire.

The variables characterising the exposure should also be specified. These can include (possibly lagged) cumulative, peak or mean exposure. Although for power computations this exposure may have to be considered as a yes/no variable, the main message cannot be that simple. Established risk factors have to be included in the final model in any case, as long as an association of it with the exposure of interest has to be assumed. On the other hand, although all measured confounders must be tested, the list of presumed confounders to be included in the main model cannot be finalized at the planning stage as they depend to a certain extent on the data themselves. E.g. a presumed confounder may not be associated with either disease or exposure of interest, in which case its inclusion is wasteful.

Practical Issues 12.7

Fund Raising 12.7.1

Technically, the expenses must be specified as regular salaries (paid by the parent organization), salaries for temporary staff, durable equipment, travel expenses, consumables (mailing, telecommunication, office material, laboratory), fees, consultants.

It is also important to adjust the flow of payment to fit the respective stages of the project. If funding comes from a number of different sources, the tentative share of each one must be specified. Experience has shown that the situation is easier to handle if only one funding agency is involved, but very large projects may require more than one source of funding (Hernberg 1992, pp 189f).

If all real costs are considered, it can be shocking to realize how expensive the project turns out to be. At this stage, the researcher must learn how to set priorities and to lower the level of ambition.

12.7.2 Data Management

Data management is a crucial step in planning a study especially with multi-centre studies. A first task in planning the data management is to identify the different data tables which are to be set up and to plan their structure and their linkage. Added complications in this context occur if the ethical requirements imply that all names (and all direct identifiers except an anonymous study identifier) are to be kept separately from some or all data tables or even have to be completely erased.

If a protocol implies merging with an external file, be it an administrative file or a file containing causes of death or diseases as is often done in historical cohort studies, a data table must be set up containing only the minimal information for merging.

The data management of prospective cohorts and other studies where subjects have to be recruited individually is especially challenging as there are several dimensions in the data. It is nearly unavoidable in such studies to keep a separate data table for the management of all contacts. In this file for every study subject all letters sent and received, all telephone contacts, all data sets or information received should be traced. Only with such a (complex) data management structure one can quickly identify non-responders and follow them up. If a subject has moved for instance, it is easier to get the new address if this change is recent. It is also important to avoid sending repeated letters to deceased subjects. On the other hand, it is important to know at each moment if a given subject is due for another contact and whether she/he has reacted to the last mailing. This implies also that this file must be able to generate the correct letter for mailing dependent on the subject's status (questionnaire to be sent, questionnaire received, lost to follow-up, pending questionnaire, etc.). The actual data received must be kept separately from this first data base.

It is also important when planning a longitudinal study to clearly identify and label the variables containing the same information collected at different time points. If these variables have the same names, they are at risk to be overwritten. Careful a priori structuring of the data base makes life much easier when creating a data file for statistical analysis. Owing to the longer time scale of such studies, the documentation is particularly important as it is not guaranteed that the same data managers will handle the data throughout the study.

Other issues to be planned carefully are data entry and coding. Three main options exist for data entry. A first option is to input data directly when interviewing the study participants. This is mainly an option with telephone interviews; using laptops in face-to-face interview may disturb the interviewee. An important aspect with direct input is the layout of the screen; the interviewer must follow the questions she/he has to ask without being disturbed by computer problems. Issues like skip patterns, the ability to easily correct already entered data, toggling between keyboard and mouse, online detection of invalid codes or inconsistencies between different data items and so on are to be programmed carefully and to be tested in real settings. A widely used computer program is the freeware Epi-Info available from the division of public health surveillance and informatics of the

Centers of Disease Control, Atlanta (http://www.cdc.gov/epiinfo). However, it has some limitations making it unsuitable for complex large-scale studies. A large range of commercial software exists, each with its own advantages. A second option is to obtain data through paper questionnaires and input the data later. In this situation a double entry is to be recommended whenever possible, especially if the questionnaires were self-administered. The data should be entered as they are, but the data problems (errors, inconsistencies) should be documented in a log-file to be treated as soon as possible. A final option is to obtain machine-readable questionnaires and apply an automated data entry based on scanning the documents and a character recognition software. Such an approach is difficult and requires careful preparation. It has been implemented in actual large-scale studies, notably in the French part of the EPIC study (e.g. Clavel-Chapelon et al. 2002). Data coding, i.e. transforming textual information (examples are places of residence, jobs, tasks or food) in a closed list of items is also an aspect to be planned and tested. It relies often on specialized knowledge. The closed list of items and coding rules must be set up before starting the data collection. Details on the construction of instruments are given in Chap. I.10 of this handbook.

A type of data that deserves a specific discussion with respect to its management are exposure measurements since they often pertain to exposure groups rather than individuals, i.e. they characterize specific circumstances (cf. Chap. I.10 of this handbook). These circumstances include e.g. measurements of air pollution in certain areas, households or occupational tasks. The main problem with these data is to be able to link them to the subject data. It is important to include the same items (i.e. labels of the exposure group) in the exposure measurement data base as in the subjects' data base. A further complication arises when some of these exposure measurements are on individuals (e.g. exposure measurement at the workplace or in households), including subjects from the epidemiological data base. These measurements characterize both the exposure group and the specific individual. Both links must then be clearly identified from the start.

Finally a log-book of all data management tasks and files should be kept. This can be part of the overall study diary.

Quality Assurance 12.7.3

Industrial standards for quality assurance and quality management are set down in the ISO 9000 series of standards (http://www.iso.ch). Their application to epidemiological studies is not straightforward given that these standards (see Moulin et al. 1998a) are geared towards customer satisfaction and that the customers of epidemiology are not easily defined. The main ideas behind these standards are however useful.

The main principles as they apply to epidemiology are the following (although a quality assurance specialist might disagree). Write up in detail what you intend to do and document what you did. Try to be proactive in thinking of what can go wrong and plan accordingly. Set up means by which you can detect any problem

as early as possible and by which you can correct your procedures accordingly. Document all changes in procedures. See also Chap. I.13 of this handbook.

We already insisted on the necessity of a detailed protocol. A protocol may furthermore be complemented by one or several standard operating procedures compiled in an mannual of operations (see below) describing the actual work to do (cook book). One main point in being proactive is to prepare the data collection in as much detail as possible. Details with respect to material conditions, e.g. hardware, software, office and storage room need to be considered in advance. Training of the data collection staff is a key to good quality data. This training should be done using the tools to be used and if possible in the setting in which the actual data collection will be conducted. It is also an important point to acknowledge that there will be non-responders and to plan how to get minimal information on those subjects from the beginning. In order to monitor errors as they arise and to be able to correct the procedures accordingly, data entry should be concurrent to the data collection and the data control and validation be done as early as possible.

Finally all changes in the protocol and in the operations manual after the start of the study must be clearly documented.

12.7.4 Study Conduct: Manual of Operations

As mentioned in the section on the study protocol, the collection of material and data as well as the methods and procedures must be described in detail. This can be done in a separate document: the manual of operations (cf. Chap. I.13 of this handbook).

The eligibility criteria must be defined in cohort studies in terms of minimum exposure, calendar time of exposure, whether or not other exposures are allowed, and so forth. If a case-control design is adopted, the eligibility criteria for both cases and controls must be well defined. For example, what histological type of cancer will be included, how will the diagnosis be confirmed? Are the controls indeed representative of the study base? Is the study a hospital-based study or a register linkage study? Who will collect the data? These questions are only examples of how detailed the description of the methods have to be. Each project has its own list of questions, so the illustration of all problems that may arise is not possible here.

The measurement methods and procedures should be described in detail. Will the indicators of disease or exposure be good measures? For example, is today's blood level of DDT a good measure of long-term exposure? Is a specially designed symptom questionnaire specific and sensitive enough to measure the neurotoxic effects of solvents? How will the interviewers be trained? Will there be a panel of radiologists for reading and interpreting the radiographs? The control and measurement of confounding should also be discussed and presented in the light of the scientific and technical background.

Time Line 12.7.5

The time schedule concerns the sequence and interdependency of different operational tasks and resources. It is good practice to outline the tasks and subtasks at the design stage and to plan their time flow, as well as the necessary resources required for each. Once it has been decided how many subjects will be included and what methods of examination will be used, the time needed can be estimated rather accurately. At the planning stage, one should make sure that statistical and computerizing assistance will be available when needed. The researcher must also stick to his/her original schedule, to avoid disrupting the consultants' time scheme. One should realize that the first data analysis will usually result in further analyses. Writing a manuscript takes time. Unexpected practical matters almost always disrupt the original time schedule. Enough time must be reserved for all these considerations. Hernberg (1992) recommends to make an allowance of half a year or more for unexpected complications.

The organisation of the time-line and the corresponding resources are best planned using project management tools like Gantt (after the method developed by Charles Gantt in 1917) and PERT (Program Evaluation and Review Technique) charts (see Figs. 12.2 and 12.3 for a simple fictitious example). Basically these tools decompose a project in elementary tasks with certain (possibly varying) durations and the corresponding needed human resources and the precedence of these tasks. Figure 12.2 presents the different tasks with arrows indicating which tasks must be terminated before the next task can start (e.g. the data collection can only start when the study has been approved by the ethics committee and when the staff has been trained). From this information, critical tasks (dark bars in Fig. 12.3) are identified, which if delayed will delay the whole study. Non-critical tasks such as preparation of data entry that may depend on completion of other tasks as for instance preparation of questionnaire can be delayed. However, non-critical tasks must be completed by the start of other work packages (here: data entry). This is indicated by the light bars in Fig. 12.3. For details of such tools see for instance Modell (1996). A number of sharewares easily available through the internet provides the software to draw these charts.

Project Diary 12.7.6

A project diary is a necessary component and should trace all aspects of the project. These aspects include the possible changes in the protocol, the data collection, and the data processing. Such a diary not only helps the investigators to keep track of the scientific realization of the operational plan of the project, but also of its administrative and economic aspects.

Changes in the original protocol may either be dictated by external circumstances but also possibly by scientific reasons, e.g. if evidence arises during the study concerning a potentially important confounder.

Documentation of data collection should not only monitor its advancement but all potentially important events. It is a particularly important issue to docu-

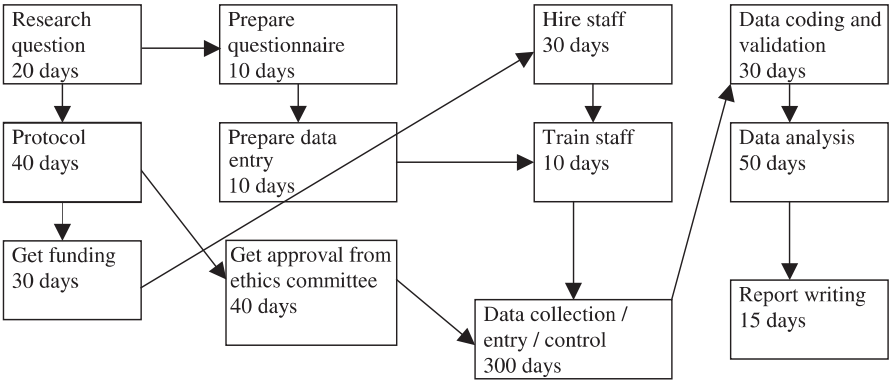

Figure 12.2. Simple fictitious example of a PERT chart for planning a study

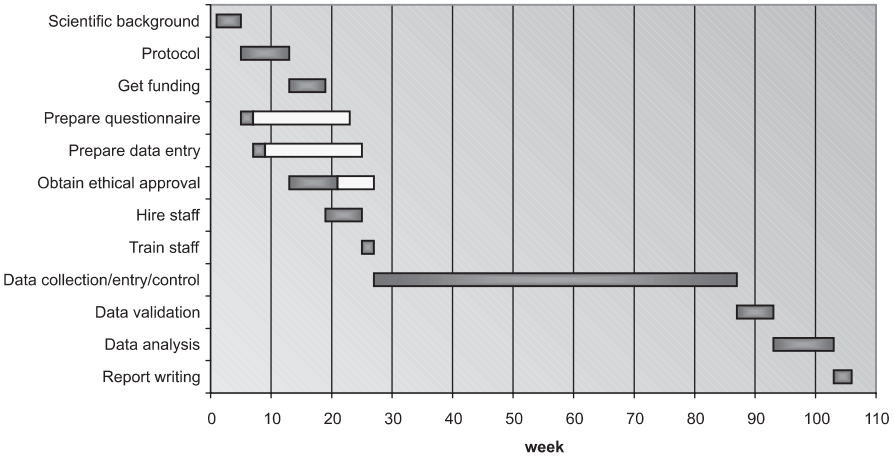

Figure 12.3. Gantt chart for planning a study

ment when and why some data could not be collected, as for instance the number of controls contacted before a control accepted to participate. It should also be documented if some technicians are on sick leave and are replaced, if some measurement instrument or computer fails and is replaced or if, at some stage, the data collection is less than optimal because of human failure. Dates and results of instrument calibration should be recorded. If some highly unlikely measurement is observed and it is redone, this information should be consigned in the diary. When some issues arise from the quality control steps (see Chap. I.13) this should also be mentioned.

Electronic data handling e.g. names of raw data files, computer transfers, merging of files, recoding, and finally all statistical data processing should all be documented including analyses which proved to be dead ends.

The project diary along with the protocol and the raw data are the central pieces of evidence with which the study can be replicated and be re-analysed, if needed.

Ethical Aspects 12.7.7

The ethical aspects are diverse and have been well covered in Chap. IV.7 of this handbook. General ethical rules for all biomedical research have been set down in the Helsinki declaration of the World Medical Organization (1996). Table 12.2 shows its central principles.

Table 12.2. General ethical principles as laid out in the declaration of Helsinki

1.	Doing good (beneficience)
2.	Not harming (nonmaleficience)
3.	Respecting persons (autonomy)
4.	Distributing goods and evils fairly (justice)

The following aspects are more specific to epidemiological research. The first ethical requirement is the respect of the data protection laws that are specific to each country. This may even require to erase part of the data, usually the personal identifiers, after a certain lapse of time. The second is the requirement that all study participants be informed of the objectives of the study, what precise medical examinations will be done and what is their purpose. The full informed consent of the participants is obligatory, at least if invasive procedures are involved. The results of each individual investigation should be made available to each study participant as well as a summary of the overall results of the study. On the contrary, confidentiality must always be adhered to and no individual results can be revealed to outsiders unless otherwise specified by law. A medical examination which might detect a hidden disease (e.g. an X-ray or a scanner might detect a lung tumor although the study was focused on pneumoconiosis) must always be medically screened before the final processing of the data in order to take the necessary medical steps. Similarly, if in the course of a prospective study it becomes evident that individual exposure levels exceed safe limits, the researcher must take the initiative to try to remove the subject from the hazardous exposure. Sensitive items in questionnaires should only be included if they are absolutely necessary. A final aspect is that the results of the study must always be made available to the community. Failing to publish the results means that the examinees have been abused and the funding has been wasted. In doing so, the interpretation of the results must be objective including a discussion of all relevant literature and all possible validity problems as well as alternative explanations of the findings.

Scientific Collaborations and Multi-centre Studies 12.7.8

An epidemiological study usually requires collaborations of the principal researcher with scientists from several fields which of course include epidemiology and usually a specialized medical field and statistics but may also include, depending on the study, genetics, molecular biology, microbiology, chemistry, industrial

hygiene, psychology, sociology. Another type of collaboration occurs if a multi-centre study is set up in which several epidemiologists combine resources. Advantages and possible pitfalls have already been mentioned and the standardization of the data collection is then a major issue. The data analysis may be centralized or decentralized. For instance in multi-centre studies organized by the International Agency for Research on Cancer (IARC), the national centres may usually publish their data and/or they are in charge of a pre-defined specific analysis of a particular topic.

In any type of collaboration the respective responsibility and role of each collaborator or collaborating centre as well as the resources allocated to the project should be clear from the start so that no false expectations arise. Provisions should be made for possibly divergent interests between study partners.

12.7.9 Publication

Epidemiological research is of interest not only to epidemiologists, but also to decision makers, funding agencies, and also to the general public. The results should be published in a form and language that they are understandable to the different target groups. This often requires two or more levels of reporting, one scientific, one popular and maybe even one press release. It is worthwhile to plan responsibilities for each aspect in advance and how the information will be dispersed. Those taking part in a study have the right to know not only their personal results but also, at least in general terms, the outcome of the whole study, especially when medical examinations are involved. The correct timing of the sequence of information delivery is important. First, those examined should be informed of their own results, then summary results should be given to funding agencies, and only afterwards to the news media. Ideally, a peer-reviewed scientific article should have appeared or at least have been accepted for publication before informing the news media. However, the scientific publishing procedure is so slow that it sometimes may be unethical to withhold urgent results from the public that long.

A large project usually gives rise to several scientific publications, and it may be useful to outline their topics in advance. At the planning stage, it is advisable to agree within the team on who will be responsible, i.e., the first author, of what, and whose names should be listed as co-authors. It is usually not possible to decide the order of names at this stage, because each team member's input to the intellectual process can be judged only after the project has been successfully completed. Guidelines as to who should be considered author of a publication are included in the Vancouver guidelines (International Committee of Medical Journal Editors 2004).

12.8 Conclusions

Careful planning is a key in the successful completion of an epidemiological study. This planning should be based on up to date scientific knowledge and awareness of all possible pitfalls inherent in epidemiological studies. It should cover all aspects from study base definition, precise design used, statistical power, control of confounding, precise data collection and exposure measurement methods to quality control, statistical methods, collaborations, dissemination of study results, and ethical issues. All these issues should be written down in a study protocol which is then a "study-bible" from which the quality of the study can be assessed.

Acknowledgements. We acknowledge the important contribution of Dr. Pascal Guenel, INSERM U170, who drafted a preliminary version of this text.

References

Ahrens W, Jöckel KH, Brochard P, Bolm-Audorff U, Grossgarten K, Iwatsubo Y, Orlowski E, Pohlabeln H, Berrino F (1993) Retrospective assessment of asbestos exposure–I. Case-control analysis in a study of lung cancer: efficiency of job-specific questionnaires and job exposure matrices. Int J Epidemiol 22 Suppl 2:S83–95

Andersen LF, Bere E, Kolbjornsen N, Klepp KI (2004) Validity and reproducibility of self-reported intake of fruit and vegetable among 6th graders. Eur J Clin Nutr 58:771–777

Armstrong BK, White E, Saracci R (1994) Principles of exposure measurement in epidemiology. Oxford University Press, Lyon

Boeing H, Korfmann A, Bergmann MM (1999) Recruitment procedures of EPIC-Germany. European investigation into cancer and nutrition. Ann Nutr Metab 43:205–215

Bogers RP, Van Assema P, Kester AD, Westerterp KR, Dagnelie PC (2004) Reproducibility, validity, and responsiveness to change of a short questionnaire for measuring fruit and vegetable intake. Am J Epidemiol 159:900–909

Breslow NE, Day NE (1981) Statistical methods in cancer research, vol I-The analysis of case-control studies. IARC, Lyon

Breslow NE, Day NE (1987) Statistical methods in cancer research, vol II-The design and analysis of cohort studies. IARC, Lyon

Breslow NE, Chatterjee N (1999) Design and analysis of two-phase studies with binary outcomes applied to Wilms tumor prognosis. Appl Stat 48:457–468

Cox DR, Wermuth N (1996) Multivariate dependencies: models, analysis and interpretation. Chapman & Hall, London

Clavel-Chapelon F, E3N-EPIC Group (2002) Differential effects of reproductive factors on the risk of pre- and postmenopausal breast cancer. Results from a large cohort of French women. Br J Cancer 86:723–727

Flesch-Janys D, Steindorf K, Gurn P, Becher H (1998) Estimation of the cumulated exposure to polychlorinated dibenzo-p-dioxins/furans and standardized mortality ratio analysis of cancer mortality by dose in an occupationally exposed cohort. Environ Health Perspect 106 Suppl 2:655–662

Garfinkel L (1985) Selection, follow-up, and analysis in the American Cancer Society prospective studies. NIH Publication 85–2713, National Cancer Institute Monograph 67. Government printing office, Washington D.C., pp 49–52

Goebel A, Netal S, Schedel R, Sprotte G (2002) Human pooled immunoglobulin in the treatment of chronic pain syndromes. Pain Med 3:119–127

Goldberg M, Chastang J-F, Leclerc A, Zins M, Bonenfant S, Bugel I, Kaniewski N, Schmaus A, Niedhammer I, Piciotti M, Chevalier A, Godard C, Imbernon E (2001) Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target dpopulation. Am J Epidemiol 154:373–384

Greenland S, Robins J (1994) Invited commentary: ecologic studies-biases, misconceptions, and counteraxamples. Am J Epidemiol 139:747–760

Hernberg S (1992) Introduction to occupational epidemiology. Lewis Publishers, Chelsea, Michigan

International Committee of Medical Journal Editors (ICMJE) (2004) Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication. Haematologica 89(3):264

Kaslow RA, Ostrow DG, Detels R,Phair JP, Polk BF, Rinaldo CR Jr (1987) The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol 126: 310–318

Kromhout H, Symanski E, Rappaport SM (1994) A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Ann Occup Hyg 37:253–270

Liu K, Stamler J, Dyer A, McKeever J, McKeever P (1978) Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis 31:319–418

Miettinen OS (1985) Theoretical epidemiology: Principles of occurrence research in medicine. John Wiley & Sons, New-York

Modell ME (1996) A professional's guide to systems analysis, 2nd edn. McGraw Hill, New York

Moulin JJ, Clavel T, Chouaniere D, Massin N, Wild P (1998a) Implementation of ISO 9002 for research in occupational epidemiology. Accred Qual Assur 3:488–496

Moulin JJ, Wild P, Romazzini S, Lasfargues G, Perdrix A, Peltier A, Bozec C, Duguerry P, Pellet F (1998b) Lung cancer risk in hard metal workers. Am J Epidemiol 148:241–248

Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press, Cambridge

Rothman KJ, Greenland S (1998) Modern epidemiology. Lippincott-Raven, Philadelphia

Rouch I, Wild P, Fontana JM, Chouaniere D (2003) Evaluation of the French version of EUROQUEST: a questionnaire for neurotoxic symptoms. Neurotoxicology 24:541–546

Rylander R (2002) Endotoxin in the environment-exposure and effects. J Endotox Res 8:241–252

Sauleau EA, Wild P, Hours M, Leplay A, Bergeret AR (2003) Comparison of measurement strategies for prospective occupational epidemiology. Ann Occup Hyg 47:101–110

Savitz DA (2003) Interpreting epidemiologic evidence. Strategies for study design and analysis. Oxford University Press, New York

Wacholder S, McLaughlin JK, Silverman DT, Mandel JS (1992a) Selection of controls in case-control studies I: Principles. Am J Epidemiol 135:1019–1028

Wacholder S, Silverman DT, McLaughlin JK, Mandel JS (1992b) Selection of controls in case-control studies II: Types of controls. Am J Epidemiol 135:1029–1041

Wacholder S, Silverman DT, McLaughlin JK, Mandel JS (1992c) Selection of controls in case-control studies III: Design options. Am J Epidemiol 135:1042–1050

Waksberg J (1978) Sampling methods for random digit dialing. J Am Stat Assoc 73:40–46

Wild P, Moulin JJ, Ley FX, Schaffer P (1995a) Mortality from cardiovascular diseases among potash miners exposed to heat. Epidemiology 6:243–247

Wild P, Refregier M, Auburtin G, Carton B, Moulin JJ (1995b) Survey of the respiratory health of the workers of a talc producing factory. Occup Environ Med 52:470–477

Wolff MS, Toniolo PG, Lee EW, Rivera M, Dubin N (1993) Blood levels of organochlorine residues and risk of breast cancer. J Natl Cancer Inst 85:648–652

World Medical Organization (1996) Declaration of Helsinki. Br Med J 313:1448–1449