A Comprehensive Clinical Epidemiological Theory Based On the Concept of the Source Person-Time and Four Distinct Study Stages

Gunnar Steineck, Philip H. Kass and Anders Ahlbom

From the Department of Oncology and Clinical Epidemiology, Oncological Centre, Radiumhemmet, Karolinska Institutet, Stockholm, Sweden (G. Steineck), the Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, California, USA (P.H. Kass), and the Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden (A. Ahlbom)

Correspondence to: Professor Gunnar Steineck, Department of Oncology and Clinical Epidemiology, Oncological Centre, Radiumhemmet, Karolinska Institutet, Stockholm, Sweden

Acta Oncologica Vol 37, No. 1, pp. 15–23, 1998

The medical community is forced to accelerate the move from opinion-based to evidence-based medicine, that is, to aim at basing all caring and clinical practice on empiri. A clear-cut epistemology would facilitate this process. In this article we present a comprehensive clinical epidemiological theory which can be used for validity issues in caring science, quality of life research, controlled clinical trials and compilations of uncontrolled evidence. The theory is based on four distinct stages that can be identified in a study, whereof the first is demarcation of the source person-time. A source person-time ('study base') can be identified for any study in all disciplines, giving an argument for using this concept as the common reference point for validity issues. Apart from identifying the source person-time, recovery of the actually observed person-time, collection of data and calculation of an ('adjusted') effect parameter (e.g., incidence ratio) are additional stages of a study. When the source person-time is demarcated confounding is introduced, when the actually observed person-time is recovered misrepresentation, in the third stage misclassification and in the fourth analytical alteration of the parameter of effect. The concept of the source person-time can, in addition, link examination of validity in caring and clinical sciences to observational studies, thereby allowing the field to benefit from all theoretical achievements for preventing, handling and understanding systematic errors developed in epidemiology. We conclude it is possible to implement a common terminology of validity for all caring and medical sciences. Drawing causal inferences in these disciplines is not mechanical, it can never, for example, be done with statistical inference. Establishing a causal relation always includes an assessment of the magnitude and direction of systematic errors influencing the adjusted effect parameter. From the presented epistemology it follows that differences in validity from a case history to a large randomized, placebo-controlled and double-blinded study are quantitative rather than qualitative. This realization in turn opens up for a more refined discussion of when a decision is evidence-based rather than opinion-based.

Received 30 January 1996
Accepted 17 June 1997

The change from opinion-based to evidence-based medicine imposes new challenges for the medical community. It is accelerated not only by internal insights, but also by an increasing external pressure. Clinical decisions are disputed by organizations providing financial resources; for example, insurance companies and governmental agencies, official authorities promote scientific revisions of medical technology and international networks such as the Cochrane collaboration are formed in order to deal with meta-analyses of controlled trials. The fact that medical practice varies geographically in a way that seemingly cannot be explained by differences in occurrence of disease is sometimes used as an argument for the need of a refined medical decision making.

A great many treatment strategies lack documentation that fulfil the standards that regulatory offices require for approval of single pharmaceutical substances. One example is the choice of approach to a patient with localized prostate cancer which largely, if not entirely, has to be based on uncontrolled evidence. Recognizing this fact, available patient series, observed independently of one another and including one treatment strategy only, have been compiled by three different groups (1–3). Methods for summarizing data and handling validity issues, were completely different in the three articles, which highlights that there is no common underlying epistemology for such work.

Epidemiologists rely primarily on observational data when they search for risk factors for disease. This has

 ISSN 0284-186X

forced the discipline to devote considerable effort into seeking to understand when observed associations can be attributed to systematic errors and when a causal inference can be drawn. In the 1970s Sackett disclosed 35 different kinds of bias in observational studies (4), and through this work most of the conceptually different threats to validity in a study were identified. Since then, attempts to identify similar types of error, that is, to collapse the systematic errors into subcategories of bias, have been made (5, 6). They are motivated by an enhanced application of theoretical and empirical achievements concerning validity, apart from pedagogical gains.

A valid clinical study precludes systematic errors by randomization, a control-arm requiring placebo and blinding of patients and investigators. Mainly single pharmaceutical substances can be evaluated with such a sharp tool. Placebo control (sham treatment) for surgery, external beam radiation therapy or combined treatments such as adjuvant multidrug cytostatic regimens are impractical and may even be unethcial. Blinding may be jeopardized by easily recognized schemes of composite treatments, such as surgery followed by cytostatic therapy, or side effects of the intervention. Finite resources, a limited number of patients, the time needed for follow-up of slow-growing tumors further compromise the possibility of obtaining data with optimal validity. Clinical studies with a substantial deviation from being randomized, placebo-controlled and double-blinded tend to resemble the observational settings found in epidemiology, indicating that the methods developed for risk factor research are pertinent.

The move forward to evidence-based medicine is borne by improved logistics to collect more valid data than before and a theoretical refinement for better design of the studies as well as an enhanced understanding of the data at hand. We have previously shown that the concept of the study base, or source person-time, can be used in defining and subcategorizing bias in observational studies (7). Here we refine this work and develop a clinical epidemiological theory, recognizing that the source person-time can be used to link the epistemology of epidemiology to discussions of validity in caring science, quality of life research, controlled trials and compilations of uncontrolled evidence.

SOURCE PERSON-TIME

The source person-time comprises subjects and time and is a defined part of the historic reality. One example from epidemiology is a cohort of twins in Sweden who completed a questionnaire in 1967 and were followed-up until 31 December 1985 (8). Another example is all subjects, 50–75 years of age, born in Sweden, and living in Stockholm for at least one month between 1 January 1985 and 31 December 1987 and observed when living in Stockholm during that time interval. This was used as the base for a case-control study (9). The concept of a source person-time that acts as a progenitor of study subjects was originally described as a 'study base' by Miettinen (10) and proceeded from a more simplistic view of study design by recognizing that the base of a study is individuals that contribute time at risk (person-time). That is, the concept of the source person-time acknowledges the essential feature of the time dimension in medicine.

An example of a source person-time in caring science is a cohort of women in Sweden all of whom were delivered of a stillborn child in 1991, observed in 1991–1994. In quality of life research, all men living in Stockholm with a new diagnosis of prostate cancer occurring in 1991 and a matched sample of men of the same age, observed during two weeks in October 1994, is another example of a source person-time (11). An example of a clinical study of randomly allocated treatment strategies is the observation of 110 women who were randomized on a protocol and observed during four chemotherapy cycles (12). Finally, the source person-time in a compilation of uncontrolled evidence may be all patients with localized prostate cancer who are included in a publication in any issue of a journal from 1980 onwards and observed until death or the end of follow-up, whichever is appropriate (1).

BIAS, DEFINITION AND SUBCATEGORIZATION

A clinical study investigates whether there is a causal relation between two factors; for example if a treatment leads to cure (elimination of disease). An example from caring sciences is if the possession of tokens of remembrance of the dead child predicts anxiety three to four years after a stillbirth (11). In quality of life research one may be interested in whether external beam radiation therapy influences sexual function (13), and a clinical question may be whether corticosteroids, given to prevent acute nausea, give a backlash and increase delayed emesis 2 to 14 days after the chemotherapy course. Finally, the question for a meta-analysis may be whether radical prostatectomy in localized prostate cancer influences survival.

Any study investigating whether a causal relation exists for any of the above-given relations can be separated into four distinct stages, each influencing validity. The study always starts with a demarcation of the source person-time. Generally, all the person-time in the source person-time is not available for collection of data, so the second step is identification of the actually observed person-time ('actual study population') from which data can be collected. The third stage of a study is collection of data, and in the fourth stage an adjusted parameter of effect is calculated.

Unfortunately, all empirical studies in medicine tend to deviate from a perfect experiment. For any study, whether randomized or non-randomized, placebo-controlled, dou-

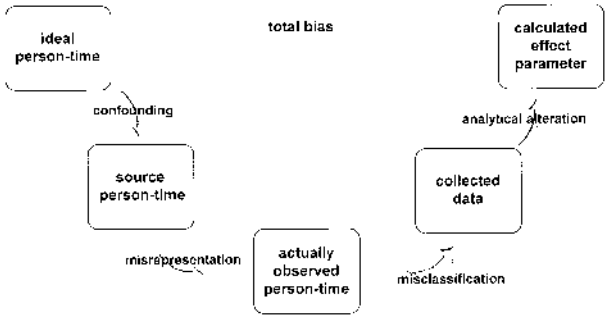

Figure 1. The four stages of a clinical study and the corresponding subcategories of bias.

ble-blinded or not, each stage defined above may introduce a systematic error that distorts the findings away from the theoretically ideal situation. In the first stage the demarcated person-time may differ from a perfect experiment so that outcome varies over the categories of the treatment (factor of interest). In the second stage, when person-time is sampled (case-control study), or when loss to follow-up perhaps occurs, the actually observed person-time may not be representative of the source person-time. In the third stage, the obtained data may not accurately reflect the situation in the actually observed person-time from which it was retrieved. Finally, the calculated effect parameter may, whether or not consciously by the investigator, deviate from the unadjusted comparison in the data.

The four types of systematic error introduced for each of the four distinct stages of a study we cite as confounding, misrepresentation, misclassification and analytical alteration of the effect parameter (see Fig. 1, Table 2). Confounding is the bias introduced when the investigator defines the source person-time. Misrepresentation may arise when the actually observed person-time is formed. Misclassification is introduced during the measurements, that is when data are collected from the actually observed person-time. Analytical alteration of the effect parameter is done in the analysis. As defined, the categories are mutually exclusive but incorporate all sources of error in a study, except for some special situations where a reversed causality between the dependent and independent variable may operate.

Total bias is the sum of confounding, misrepresentation, misclassification and analytical alteration of the effect parameter. Note that this is not an additive function of the absolute values of the subcategories, that is not the sum of the absolute values of confounding, misrepresentation, misclassification and analytical alteration. Some systematic errors can have opposite effects that tend, in part, to cancel each other (Table 3). Specifically, in the analysis modeling or stratification is done in order to diminish bias, the investigator consciously tries to remove some of the bias introduced by confounding, misrepresentation or misclassification.

CONFOUNDING

In a perfect experiment there is no difference between the two groups that are being compared, had the therapy (factor of interest) not been instituted. Such a theoretical situation occurs if a very large number of subjects are randomly allocated to two groups and thereafter observed for one year. Another imagined ideal situation would be if we could simultaneously observe a patient's response to treatment and non-treatment, a hypothetical setting that would hold all extraneous factors constant.

Table 1

Hypothetical examples of a source person-time, the corresponding actually observed person-time and the source of misrepresentation when the source person-time is not identical with the actually observed person-time

Source person-time	Actually observed person-time	Source of misrepresentation
All subjects randomized in a specific study and the planned follow-up until death	The follow-up that actually was possible among the randomized patients	Loss to follow-up
All subjects with a diagnosis of prostate cancer in Stockholm 1991, observed in a cross-section on one date in 1994	All subjects actually responding to a posted questionnaire sent to identified subjects with prostate cancer observed the date the questionnaire is answered	Diagnosed but non-identified patients Non-responding patients (who can be regarded as a loss to follow-up)
All subjects born during 1911–1945 living in Stockholm for at least one month from 1985 to 1987	All cases of bladder cancer occurring in the source person-time that responded to a questionnaire, selected controls that responded to a questionnaire	Invalid selection of controls. Diagnosed but non-identified cases Non-responding cases Non-responding controls
All patients with localized prostate cancer observed during 1980–1990	Published patients with localized prostate cancer observed during 1980–1990	Selective processes in the choice of observed patients for publication
All published patient series with localized prostate cancer observed during 1980–1990	Half of the published patient series with localized prostate cancer observed during 1980–1990	Selective processes in identifying articles for a compilation

Table 2

Sources of systematic error in a clinical study

Label of the categories of systematic error	Sources of systematic error in a clinical study of treatment efficacy
Confounding (Discrepancy between the source person-time and the ideal person-time)	1. An imbalanced distribution of prognostic factors between the treatment groups 2. An imbalanced distribution of extraneous factors, e.g., a concurrent treatment, affecting outcome between the treatment groups 3. An imbalanced observation of parts of a hazard curve that varies over time between the treatment groups 4. The outcome influences the definition of the source person-time
Misrepresentation (Discrepancy between the actually observed person-time and the source person-time)	1. Part of the source person-time is lost before measurement of outcome or person-time –Loss to follow up –Unrepresentative sample of person-time ('controls') events of outcome 2. Observed person-time is not published 3. Published person-time is not incorporated in a compilation
Misclassification (Discrepancy between the date and the actually observed person-time)	1. Measurement error of events of outcome 2. Measurement error of person-time
Analytical alteration of effect parameter (Discrepancy between the adjusted effect parameter and the data)	1. Errors in calculation of the effect parameter 2. Changes in the effect parameter after adjustment for covariates: Diminished bias: adjustment for appropriate covariates. Increased bias: adjustment for inappropriate covariates, covariate misclassification, inadequate statistical techniques
Some means of influencing the systematic errors in a controlled study*	**Some means of influencing the systematic errors in an observational study**
Random allocation of treatment	Identification of prognostic factors of outcome ('extraneous risk factors'), restriction or adjustment in analysis
Placebo control (sham treatment) in the comparison group (among non-treated patients)	Identification of concurrent treatments that may affect outcome restriction or adjustment in analysis
Observation of all patients from the point of randomization and having identical periods of follow-up between groups	Restriction to subgroups with similar time from diagnosis to start of observation and similar periods of follow-up
No exclusion of patients after randomization	Restriction to subgroups in which no exclusions were made
Securing of a complete follow-up. 'Binding' to ensure that investigator-related loss to follow-up is independent of the treatment given**	Restriction to subgroups in which a high percentage were followed up Valid choice of controls, e.g. a random sample of the source person-time Avoidance of non-response among cases and controls
Securing a high sensitivity and specificity	Publication of all observed person-time and including all published person-times in a compilation
'Binding' to assure that investigator-related misclassification is independent of the treatment given**	Securing a high sensitivity and specificity. Restriction to subgroups in which misclassification is judged to be independent of treatment given
No errors in the analysis	No errors in the analysis
Use of appropriate statistical models and proper choice of covariates for restriction, modeling or regression (adjustments)	Use of appropriate statistical models and proper choice of covariates for restriction, modeling or regression (adjustments)

* Either to diminish bias or ensure that bias will dilute the difference between treated and untreated patients.
** Bias will dilute or not affect any treatment effect, depending on whether it is measured as a ratio or difference.

If the source person-time in a specific study has a varying outcome between treated and non-treated subjects (person-time), independent of the factor of interest, it will differ from the situation in a perfect experiment. The systematic error introduced, confounding, is an issue for all clinical studies. If women with a high socioeconomic status to a greater extent collect tokens of remembrance, and if socioeconomic status is related to anxiety, this factor confounds the association between token of remembrance and anxiety. Men with cardiovascular diseases may be selected for external beam radiation therapy rather than a radical prostatectomy, and since cardiovascular disease influences well-being, this factor may confound a comparison between the two treatment strategies. By chance, more women prone to a high level of delayed emesis may be randomized to corticosteroids rather than to placebo, confounding the comparison (14). Finally, patient selection, lack of placebo and observation of different parts of a hazard curve may confound a comparison of independently collected series of prostate cancer.

It is instructional to view confounding as a result of an imbalanced distribution between groups of 1) disease char-

Table 3

*Exact definition and subcategorizing of factors influencing validity according to a scheme with distinct study stages. Rate ratio is used as the effect parameter**

Abbreviation used below	Explanation
RR (ideal person-time)	True rate ratio in a theoretically ideal person-time for a study
RR (source person-time)	True rate ratio in the source person-time
RR (actually observed person-time)	True rate ratio in the actually observed person-time
RR (data)	True (unadjusted) rate ratio inherent in the data
RR (adjusted)	Calculated and adjusted rate ratio
Label of the categories of the sources of error	**Definition of the categories of the sources of error**
Confounding	RR (source person-time)–RR (ideal person-time)
Misrepresentation	RR (actually observed person-time)–RR (source person-time)
Misclassification	RR (data)–RR (actually observed person-time)
Analytical alteration of effect parameter	RR (adjusted)–RR (data)
Total decrease in validity	Confounding + misrepresentation + misclassification + analytical alteration
Total decrease in validity	RR (adjusted)–RR(ideal person-time)

* Modified from reference 7.

acteristics (prognostic factors) affecting outcome, and 2) other factors affecting outcome such as concurrent treatments. Confounding may also occur when 3) different parts of a hazard curve which is not constant over time are observed (15). Finally, 4) a flawed definition (influence on the definition by the treatment effect studied) of the source person-time gives confounding (a 'biased base').

1. In non-randomized studies, patient selection can lead to groups that are inherently different with respect to the incidence of the outcome. For example, in the compilation of series with localized prostate cancer it was found that the proportion of poorly differentiated tumors differed between patient series (1). Degree of differentiation is a strong prognostic factor of death in prostate cancer (16), and confounding is thus introduced in a comparison between the compiled groups. In oncology, the TNM system was developed to provide a means for understanding this type of confounding.
2. Treatments other than an investigational drug or a surgical technique may differ between the groups that are compared. If, for example, treatment of metastatic bladder cancer with a multidrug cytostatic regimen (MVAC) is to be compared with historical controls, one must assess the influence of any surgery performed after MVAC to excise metastases in some patients (17). It may be surgery, not the cytostatics, that cured the patient. Another example is that when the outcome of a new surgical technique is compared with historical controls, one must consider differences in supportive care during and after the operation in the contemporary series as opposed to the historical data. Placebo control ('sham treatment') is used in controlled trials to outweigh extraneous concurrent 'treatments' that may differ between groups.
3. Non-comparability also emerges when an individual's instantaneous rate of outcome (hazard) varies with time, and when treatment groups consist of individuals whose incidence rates are observed at different times along the mentioned hazard curve. If asymptomatic subjects are screened for a certain disease, on average the diagnosis of cancer in this population will be made earlier than in a population without screening. If the outcome of, say, death of the specific cancer is compared between the screened and unscreened populations, the 'lead time' introduced may give a spuriously decreased risk of cancer-specific death in the screened population (18, 19). A similar, situation occurs when compiled series of localized prostate cancer patients, having undergone external beam radiation therapy, are compared with series with deferred therapy. In the former category a longer time may have elapsed from diagnosis of cancer to the start of treatment, which may introduce confounding in the comparison if the observation period begins with the treatment ('inverse lead time') (1).
4. It is possible for an investigator to be influenced by the outcome under investigation when he defines the source person-time, a phenomenon described as a biased base by Miettinen (10). For example, in a retrospective study conducted at a center some patients may have died of the disease under study in institutions other than the center instituting the investigation. If the source person-time is defined to include patients with follow-up at the center only, the mortality of the disease will be underestimated, and a bias will be introduced if it is compared with another group with a true complete follow-up.

One cornerstone for preventing confounding in the design of a study is random allocation of treatment. On

average, randomization will distribute prognostic factors of outcome evenly between the groups. Another cornerstone is placebo control (sham treatment), which at best evens out between groups all concurrent factors affecting outcome. When randomization is not possible, cross-over studies, with sequential comparison of exposure and non-exposure in the same individuals are sometimes conducted. Another pseudo-experimental setting is to observe a geographically determined allocation of treatment strategies between subjects. To understand the possible impact of confounding it is important to collect information on all potential confounding factors. In the analysis (*vide infra*), restriction or 'adjustment' (stratification or regression modeling) for confounding factors may diminish confounding. Confounding variables may reflect a prognostic factor, a concurrent treatment or subgroups with varying time-dependent relations, such as time from diagnosis to start of follow-up.

MISREPRESENTATION

In the process of going from the demarcated source person-time of the study to identifying the actually observed person-time, misrepresentation may be introduced. Examples are loss to follow-up, a biased sampling of controls in a case-control study and non-response among cases and controls in a case-control study built on interviews (Table 1).

Loss to follow-up can bias the results, give misrepresentation, when it leaves an actually observed person-time that differs from the source person-time concerning the relationship between the factor of interest and outcome. If women without tokens of remembrance and anxiety are more inclined to return a postal questionnaire than women with a token of remembrance and anxiety, and there is no difference in follow-up between groups for women without anxiety, the findings become distorted. In a compilation of independently collected series of patients with localized prostate cancer, a loss to follow-up related to death from prostate cancer can be detrimental if prostate cancer mortality is studied and it differs between the treatment strategies of interest (1). If controls are sampled from the source person-time (case-control study), for example to illustrate the distribution of baseline prognostic factors, bias (misrepresentation) is introduced if the sample is not representative of the source person-time concerning the factors ('exposures') of interest.

In a comparison based on a compilation of independently collected patient series, two further sources of error are of interest and they can be classified as misrepresentation. Consider a source person-time defined as 'all patients in the world *observed* at any time during 1980–1985', and an actually observed person-time of the experience actually published. A 'publication bias', including the preferences of editors, may create a difference between the actually observed person-time and this source person-time. If treated patients with an unfavorable outcome were not published, this might imply a serious error in a compilation based on published series. A second source of error in any compilation is a potential selection among the published articles. If all published articles are not included, the investigator has mainly included articles that favor, for example, a treatment effect for a certain agent and a bias (misrepresentation) is introduced in the compilation.

In the design of a study, randomization can diminish the possibility of misrepresentation if it is related to factors present at the baseline of the study. It will not prevent bias from loss to follow-up caused by the treatment itself. Blinding may make the modes of follow-up similar in the treatment groups and prevent a selective loss of the source person-time. If the loss of person-time is the same in the groups that are compared, misrepresentation will, on average, dilute or not affect the observed effect parameter.

In the analysis (*vide infra*), restriction, stratification and regression may diminish misrepresentation. A prerequisite is that a covariate is included which reflects subgroups with, for instance, varying loss to follow-up. For example, restricting a study according to a covariate that demarcates patients when there is no loss to follow-up would probably improve the validity.

MISCLASSIFICATION

In the third stage of a study, data are collected from the actually observed person-time and in this process misclassification occurs. A sensitivity or specificity of less than 100% in the measurement of the factor of interest (e.g. treatment) or outcome defines misclassification (20).

A major threat to validity in studies with uncontrolled evidence is a varying misclassification of outcome between groups. For example, women with a token of remembrance may report any specific level of anxiety differently than women without such a token. Another example is when series of patients with localized prostate cancer are compiled and local recurrence is the outcome. Among patients having undergone external beam radiation therapy, a local recurrence is determined more often by a digital rectal examination only as compared to patients having undergone radical prostatectomy (21). Among the latter, a biopsy of the prostate is more common, a method conceivably having a higher sensitivity for tumor growth than digital rectal examination only. Thus a bias due to differential misclassification between groups concerning outcome is introduced if the compiled data are used in a comparison (21).

Sensitivity in assessing outcome is defined as the proportion of subjects with a true event of outcome who are defined as having such. The comparison is thus a 'true' event, which is a theoretical construct. One way to formulate this construct is to refer to a 'conceptual scale' (10). For example, an event of recurrence of prostate cancer can theoretically be thought of as the point when a growing neoplasm in a patient reaches a total of 1 000 cells. In practice, one can never determine sensitivity and specificity

with certainty, since the true event exists only in theory and cannot be determined empirically. No 'golden standard' can be found for any variable, be it a module in a questionnaire measuring quality of life or data obtained with molecular biological techniques. Different ways of measuring a specific variable can be compared, but indirect methods have to be used to illustrate the sensitivity and specificity.

In the design of a study, randomization can diminish the possibility of misclassification that varies between groups if it is related to factors present at the baseline. It will not prevent misclassification which is related to the treatment itself. Blinding is instituted in controlled trials to ensure that investigator-related sources of misclassification are non-differential between the treatment groups. If the decreased sensitivity and specificity are the same in two compared groups (non-differential misclassification of outcome), the bias introduced will neither dilute nor affect the observed treatment effect.

In the analysis (*vide infra*) restriction, stratification or regression ('adjustments') may diminish misclassification. A prerequisite is that a covariate reflects subgroups with, say, a varying degree of sensitivity in the detection of the outcome event. For example, if the detection of an event of distant metastasis is done with a higher sensitivity among women than among men when adjuvant chemotherapy is studied, adjustment for gender will decrease the bias introduced due to misclassification.

ANALYTICAL ALTERATION OF THE EFFECT PARAMETER

The fourth and last stage of a study results in a calculated effect parameter (e.g., incidence ratio or difference, prevalence ratio or difference). In the analysis a major effort is often made to remove some of the bias introduced by confounding, misrepresentation and misclassification by adjustment for appropriate covariates. We use the term 'analytical alteration of the effect parameter' since it does not allude to an error, the magnitude of total bias may increase, but is often decreased in this stage of the study (7).

The analysis of a study serves a twofold purpose: to summarize and smooth the data for descriptive and inferential purposes (22, 23), and to attempt to remove bias introduced in the earlier stages of the study. The latter can be done by restriction, stratification or regression. Stratification and regression (modeling) are often called 'adjustment' of the crude effect parameter in observational studies. When the 'adjustment' is successful, it will diminish bias introduced earlier by confounding, misrepresentation or misclassification. The 'adjusted' effect parameter is closer to the effect parameter in an 'ideal study' (Fig. 1) than the crude or unadjusted effect parameter.

There are situations, however, when the bias is also introduced in the analysis, apart from pure mistakes. Under certain conditions, bias may increase due to adjustment for inappropriate covariates, adjustment by misclassified covariates or inadequate statistical models (24–26). This is not least important to consider when several prognostic factors are included in statistical models. When 'adjustment' is incomplete due to misclassification of the true confounding factor, or imperfect coverage of misrepresentation and misclassification, the change in the effect parameter (that is, the analytical alteration) does not completely correct for the bias previously introduced. A specific bias arises when certain statistical assumptions are violated, which has been called specification error (5).

In the design of a study, the possibility of diminishing bias in the analysis increases if covariates reflecting confounding, and possibly also misrepresentation and misclassification, are identified and arrangements are made to measure them accurately. During the analysis, use of a priori information concerning the covariates can increase the probability that appropriate variables are employed for stratification or modeling. For example, data elucidating whether a factor is a true prognostic factor for the incidence of outcome may be important.

DISCUSSION

We have presented a theoretical framework for understanding clinical research which includes a definition of the total bias as well as four exclusive subcategories of it. Our model can refine the communication and interpretation of medical information, and provides a common scheme for caring sciences, quality of life research, controlled clinical trials and compilations of uncontrolled evidence. When an error has been identified as belonging to one of the categories of bias, theoretical achievements and empirical observations of the specific category can be employed in the interpretation of findings. With clear definitions of subcategories of bias, a prerequisite for a uniform terminology in different disciplines is fulfilled.

Observational studies of risk factors for disease and clinical research providing a basis for not over- or undertreating patients are two research traditions where a merge has been discussed for quite some time now. Hill mentioned early on that a tradition comprising observational studies must always 'keep the experiment firmly in mind' (27), emphasizing the epistemological kinship between experimental and non-experimental studies in medicine. Treatment in a controlled clinical trial, or in uncontrolled evidence, corresponds to 'exposure' in etiological research.

All studies are susceptible to confounding, only in hypothetical situations is a study free from this error. Misclassification can never be fully assessed empirically, and it is feasible that most, if not all, studies include misrepresentation. The inadequacy of any empirical investigation to

avoid systematic errors has been called 'the fundamental problem of causal inference' (28) and has prompted Greenland to conclude that a causal inference can never be made with statistical inference only in observational studies (14). Statistics are used to summarize data in an effect parameter, and statistical techniques, such as stratification or modeling, are employed to diminish the systematic error in the calculated ('adjusted') effect parameter (6). Causal inference always includes a judgement concernng to what extent residual systematic errors influence the effect parameter after adjustment. Our clinical epidemiological framework comprises the mental tools to render this assessment of causality effective (29).

While the discussion of concepts of validity is an important one, harmonizing terminology is also desirable, not least for promoting a dialogue between different disciplines. Some words have different and confusing meanings in the literature, not least the term 'selection bias'. For a surgeon selecting patients for surgery, it may be natural to cite the error introduced when his case material is compared with a control group as 'selection bias' rather than confounding. Investigators responsible for a randomized trial may be concerned with a selection of patients to the trial, a phenomenon they may refer to as a 'selection bias' instead of discussing the generality of results, i.e., the presence of effect modification. 'Selection bias' is also used for misrepresentation introduced by non-response in a case-control study. Another, less important, example is that the division of systematic errors in bias and non-bias components, including confounding, is advocated by some (30, 31), but not by others (5). Furthermore, two unrelated definitions of confounding can be encountered in the literature (32) and ours is close to the 'comparability-based' definition rather than the 'collapsibility-based' one.

Misrepresentation poses an important threat to validity in all studies. Although possible in theory (vide supra), a priori knowledge can seldom be used to assess its possible size and with the data from a specific study it is rarely possible to assess the magnitude of the bias introduced, or in which direction it influences the effect parameter. Thus, it is important for the assessment of validity to distinguish confounding and misclassification from misrepresentation. A biased base, gives an error of the same unpredictable sort as misrepresentation. Confounding, apart from a biased base, is well characterized theoretically and there are well-developed methods for handling it. Since a confounding factor must be causally related to the outcome, or be a surrogate variable for such a causal factor, all knowledge at hand prior to the specific study concerning such a possible relation can be used when a study is planned and in assessing whether confounding is an issue afterwards. The magnitude of confounding depends on its strength in predicting outcome and the degree of correlation with the treatment. Moreover, the influence of the misclassification of a confounding factor has been examined. In countries which lack population-based registries, the distinction between the source person-time and accessible person-time is sometimes diffuse in specific studies, and thereby the possibility to understand whether the systematic error involved behaves like confounding or like misrepresentation.

Misclassification, in addition, has a vast literature, and the magnitude and direction of the error introduced can often be estimated with some accuracy. For example, it may be counterintuitive that a radiographic investigation missing half of the patients with metastases (50% sensitivity) does not introduce a systematic error if the effect parameter is incidence ratio.

There is a wide spectrum of designs for retrieving medical information from a case history to a large, randomized, placebo-controlled, double-blind study. Reports of single cases are used as a source of knowledge in authoritative medical journals and a formalization of such a report can start with a 'source person-time' comprising the observation of one subject for a specified time and person-time historically observed. The spectrum from the case history to the most valid study includes a compilation of independently collected patient series, quasi-experimental matching of treated patients with non-treated referents, random allocation of treatment but no placebo control or blinding, blinding in the assessment of outcome but no randomization, and so on. Recognition of the fact that these varying designs imply quantitative rather than qualitative differences in validity, and open up the possibility of a more highly developed handling of different sources of systematic error than before, may not only accelerate the evolution of knowledge and thereby the speed with which we move from an opinion-based to an evidence-based practice, in care and clinically, but also refine the judgement whether practice is evidence-based or not.

ACKNOWLEDGEMENTS

This study was supported by grants from the Swedish Cancer Foundation.

REFERENCES

1. Adolfsson J, Steineck G, Whitmore WF Jr. Recent results of management of palpable clinically localized prostate cancer. Cancer 1993; 72: 310–22.
2. Austenfeld M, Thompson IM, Middleton RG. Meta-analysis of the literature: guideline development for prostate cancer treatment. J Urol 1994; 152: 1866–9.
3. Wasson JH, Cushman CC, Bruskewitz RC, Littenberg B, Mulley AG, Wennberg JE. A structured literature review of treatment for localized prostate cancer. Arch Fam Med 1993; 2: 487–93.
4. Sackett DL. Bias in analytic research. J Chron Dis 1979; 32: 51–63.
5. Greenland S. Concepts of validity in epidemiological research. In: Holland WW, Detels R, Knox G, eds. Oxford Textbook of Public Health, Volume 2, Methods of Public Health, 2nd ed. New York: Oxford University Press, 1991; 253–70.

6. Rothman KJ. Modern epidemiology. Boston: Little, Brown, & Co, 1994.
7. Steineck G, Ahlbom A. A definition of bias founded on the concept of the study base. Epidemiol 1992; 3: 477–82.
8. Steineck G, Norell SE, Feychting M. Diet, tobacco, and urothelial cancer; a 14-year follow-up of 16 477 subjects. Acta Oncol 1988; 77: 323–7.
9. Steineck G, Hagman U, Gerhardsson M, Norell SE. Vitamin A supplements, fried foods, fat and urothelial cancer. A case-referent study in Stockholm in 1985–87. Int J Cancer 1990; 45: 1006–11.
10. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley and Sons, 1985.
11. Rådestad I, Steineck G, Nordin C, Sjögren B. Psychic complications after stillbirth in relation to memories and immediate management. Br Med J 1996; 770–80.
12. Fredrikson M, Hursti T, Furst CJ, et al. Nausea in cancer chemotherapy is inversely related to urinary cortisol excretion. Br J Cancer 1992; 65: 779–80.
13. Helgason AR, Fredrikson M, Adolfsson J, Steineck G. Decreased sexual capacity after external radiation therapy for prostate cancer impairs quality of life. Int J Radiation Oncology Biol Phys 1995; 32: 33–9.
14. Greenland S. Randomization, statistics, and causal inference. Epidemiol 1990; 1: 421–9.
15. Steineck G, Adolfsson J. Bias due to observation of different parts of a hazard curve over time. Urology 1996; 48: 105–9.
16. Chodak GW, Thisted RA, Gerber GS, et al. Results of conservative management of clinically localized prostate cancer. New Engl J Med 1994; 330: 242–8.
17. Sternberg CN, Yagoda A, Scher HI, et al. M-VAC for advanced transitional cell carcinoma of the urothelium: Efficacy, and patterns of response and relapse. Cancer 1989; 64: 2448–58.
18. Hutchinson GB, Shapiro S. Lead time gained by diagnostic screening for breast cancer. J Natl Cancer Inst 1968; 41: 665–81.
19. Feinleib M, Zelen M. Some pitfalls in the evaluation of screening programs. Arch Environ Health 1969; 19: 412–5.
20. Mertens TE. Estimating the effects of misclassification. Lancet 1993; 342: 41821.
21. Steineck G, Whitmore WF Jr, Adolfsson J. 'Local recurrence' and 'disease-free survival'; doubtful parameters when comparing non-randomized studies of prostate cancer. Scand J Urol Nephrol 1991; 138: 121–6.
22. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989; 79: 340–9.
23. Greenland S. Summarization, smoothing, and inference in epidemiologic analysis. Scand J Soc Med 1993; 21: 227–32.
24. Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol 1985; 122: 495–506.
25. Ahlbom A, Steineck G. Aspects of misclassification of confounding factors. Am J Ind Med 1991; 21: 107–12.
26. Leon DA. Failed or misleading adjustment for confounding. Lancet 1993; 342: 479–81.
27. Hill AB. Observation and experiment. New Engl J Med 1953; 248: 995–1001.
28. Holland PW. Statistics and causal inference. J Am Stat Assoc 1986; 81: 945–60.
29. Norell S. A short course in epidemiology. New York: Raven Press, 1992.
30. Sitthi-amorn C, Poshyachinda V. Bias. Lancet 1993; 342: 286–8.
31. Datta M. You cannot exclude the explanation you have not considered. Lancet 1993; 342: 345–7.
32. Kass PH, Greenland S. Conflicting definitions of confounding and their ramifications for veterinary epidemiologic research: collapsibility vs comparability. J Am Vet Med Assoc 1991; 199: 1569–73.