J Chron Dis Vol. 32, pp. 15 to 27
Pergamon Press Ltd 1979. Printed in Great Britain

THE EVOLVING CASE-CONTROL STUDY*

PHILIP COLE†‡

I. INTRODUCTION

MY ASSIGNMENT is to point out limitations in our understanding of the case-control study and to suggest areas where new knowledge should improve the technique. I address this task as a practicing cancer epidemiologist. If some of the problems I describe seem esoteric, it is due to only a thin veneer. In fact, these are real problems which hinder me every day as I try to do, and try to teach others to do, case-control studies.

During the 1950's many fine case-control studies were done, and we can learn from them as examples of the technique. And during the same decade there emerged, I believe for the first time, a series of studies of the technique *per se*. I began by reviewing these basic papers written in the 1950's by Jerome Cornfield, Harold Dorn, Nathan Mantel and William Haenszel [1–5]. The decades which have passed since these works first appeared provides great perspective. To be sure, one now sees more clearly some limitations in these writings. That is because the same researchers, among others, have reduced them. But, overwhelmingly, the study of these papers shows that epidemiologists who have done case-control studies during the past 20 yr could have stood on the shoulders of giants. (And, lest we epidemiologists lose sight of one major root of our science, we should remember that all of these men are, or were, statisticians.) These four writers had such an appreciation of problems and such clarity of thought that we still rely to a great extent on their work for our understanding of the case-control method. Indeed, most present-day criticisms of the case-control study are countered by their 20 yr-old writings.

Let us begin with two definitions of epidemiology. In the 1930's Gaylord Anderson defined epidemiology as "the science of disease occurrence" [6]. I have expanded this definition into the following: epidemiology is the science dealing with the environmental causes of diseases of humans as inferred from observations of human beings. As a practical point one should emphasize that at present most epidemiology is non-experimental. It is this feature, the use of non-experimental methods, which causes many of the difficulties in our work. (Parenthetically, this feature also relates to the great strength of epidemiology, namely, its relevance to the problems of man. Epidemiology is the study of the diseases of man in man and it is an ethical corollary of this which usually precludes experimentation.)

The restrictions imposed on epidemiologic research by its non-experimental nature are most burdensome in case-control studies. By a case-control study I mean an investigation of the exposure frequencies of at least two groups of subjects selected on the basis of their status with respect to a particular disease entity. Nearly every word would require elaboration before this description would be acceptable to more than a small minority of epidemiologists. I shall not dwell on this as several of the larger issues

*From the Unit of Epidemiology and Biostatistics, International Agency for Research on Cancer, Lyon, France. Supported by a contract (N01–C0–55195) from the International Cancer Research Data Bank, National Cancer Institute, U.S.A.
†Present address (and reprint requests): Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.
‡I am grateful to Madame J. Nectoux for compilation of some of the data presented; to Dr. John Cooper for a critical review of the manuscript; and to the Symposium participants for helpful comments.

implied by the description form the substance of this paper. However, I implied a polytomous classification as far as health is concerned, rather than the usual dichotomy of diseased and non-diseased, that is, of cases and controls, because for some disease entities (atherosclerosis, hypertension, mental illness) it may be useful or necessary to include a group intermediate to the clearly ill and the clearly healthy, if not in the data gathering, then in the analysis. Nonetheless, most discussions of case-control studies are facilitated if limited to the usual two groups and I will so restrict myself.

II. CURRENT STATUS

Probably the first true case-control study of the modern type was reported in 1926 by Lane-Claypon; it was an investigation of the role of reproductive experiences in the etiology of breast cancer [7]. By 'true', I mean that observations were made on a defined control group, as individuals, and that these observations were at least approximately equal in quantity and quality to those made on the cases. By 'modern', I mean that the study showed the special suitability of the case-control method for a difficult problem, namely, the long induction period of the chronic diseases. [There is, of course, another valuable approach to overcoming a long induction period, the non-concurrent follow-up study (also known as the retrospective cohort study and by other names). But such studies require the good luck of locating old, but pertinent, information.]

The case-control study is valuable because it permits us to see back through time, from effects back to causes. True, this is not the sequence of logic which we use in experimental research, nor in follow-up studies whether concurrent or non-concurrent. Nonetheless, we need not apologize for the case-control study as it is not backward, unnatural, or inherently flawed. Indeed, in everyday human affairs cause–effect relationships are frequently viewed in the reverse of their temporal sequence, but we have no difficulty in understanding them. However, everyday affairs usually have causal paths that are short, simple and strong. When a causal path spans decades our ordinary perceptions may not suffice. This is all the more so if the path is made fainter yet because the cause–effect relationship is weak, as is usual in chronic diseases. So we do need a special method of observation to see back through time. That special method is the case-control study. I do not contend that the case-control study is a flawless time machine. It has shortcomings, some of which we understand poorly. We can only try to perceive these and to rectify them while taking advantage of this unique research tool.

In addition to its general advantage, of seeing back through time, there are several more specific reasons for the popularity of the case-control study. There is the empiric observation that *it works*. This was exemplified by the earliest case-control studies of cigarette smoking and lung cancer in the 1950's [8, 9]. At about the same time Cornfield showed how to transform the relative exposure frequencies acquired in a case-control study into a parameter of far greater interest to public health workers, the relative incidence (also known as the odds ratio, relative risk and by other names) [1]. Later, the synthesis of Mantel and Haenszel clarified the objectives of case-control studies, systematized the issues to be confronted and also described two of the techniques now most widely used in the analysis of case-control studies [4]. It is encouraging that a review of the frequency of citations of papers which have appeared in the **Journal of the National Cancer Institute** showed the Mantel–Haenszel paper to be in sixth place, one of only two epidemiologic papers in the first 50 [10]. Moreover, its use is increasing rather than waning like that of most of the other 'top 50'. This suggests that we will see not only more, but better, case-control studies.

I decided to try to quantify the widely-held, but subjective, impression of the growing popularity of the case-control study. I enumerated the several different types of articles appearing in four medical journals in two 2-yr time periods. The time periods are 1956–1957 and 1976–1977. The journals are **The Lancet**, **The New England Journal of Medicine (NEJM)**, the **American Journal of Epidemiology** and the **Journal of Chronic Diseases**. For the two general journals all original articles were classified as reporting

a case-control study, a case series, any other epidemiologic study (usually a prevalence survey or incidence survey, rarely a follow-up study) or not an epidemiologic study. For the two specialty journals all articles were classified only as a case-control study or not. It was considered that: (1) an article is a report of an epidemiologic study if it is based on observations made to elucidate the causes of a disease of unknown but presumably environmental etiology, (2) an article is a report of a case-control study if it meets the preceding criteria and if it is based on individual persons as the unit of observations and these persons are deemed to be ill or not ill with a specified disease, (3) a case series is an aborted case-control study; there is no control group but there may be some basis for suggesting that cases have an unusual (usually high) frequency of exposure to some presumptive cause of the disease.

Results for two of the journals are shown in Fig. 1. Although **The Lancet** reduced the number of original articles published from 420 to 325/yr over this 20-yr span, the number of epidemiologic papers of all types increased nearly 2.5-fold, from 38 to 98. The number of case-control reports increased seven-fold, from 7 to 48, and the percentage of case-control reports increased nine-fold, from 0.8 to 7.4%. The percentage of 'other' epidemiologic reports increased about two-fold from 2.1 to 4.6%. Unlike **The Lancet** the **NEJM** published almost as many articles per year in the mid 1970's (195) as it had in the mid-1950's (209). The number of epidemiology papers in the **NEJM** increased both for case-control studies (from 0.7 to 3.3%) and for 'other' epidemiologic studies (from 1.2 to 5.9%). One peculiarity of the **NEJM** data is that there were only three case-control studies reported in 1977 compared to 10 in 1976. For the **American Journal of Epidemiology**: in 1956–1957, 119 articles were published and none was a case-control study. In 1976–1977, 201 articles were published (excluding two special issues) including 17 (8.5%) reports of case-control studies. For the **Journal of Chronic Diseases**: in 1956–1957, 179 articles were published including one (0.6%) case-control study. In 1976–1977, 124 articles were published of which two (1.6%) were case-control studies.

In summary, both general medical journals now publish more epidemiologic studies. **The Lancet** has a propensity to publish case-control studies, the **NEJM** has not. It seems the British are determined to hold the lead they have always had over Americans in epidemiology; once again they are showing us what we will be doing in the future. In any event, it is clear that the *number* of case-control studies published in these journals has increased four- to seven-fold over the 20-yr period. With respect to the specialty journals, the data for the **American Journal of Epidemiology** support the concept of an increase in case-control studies both in terms of numbers and percentage of such

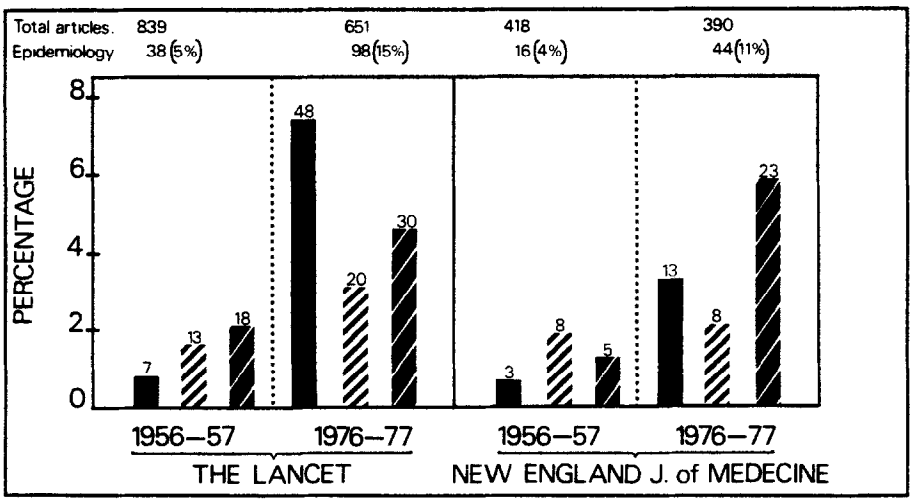

FIG. 1. Percentage of total published articles that are reports of case-control (first bar), case series (second bar) or other epidemiologic studies, for two general medical journals and two time periods. (Number of articles atop each bar.)

papers published but the **Journal of Chronic Diseases** does not. Thus, three of the four journals show that case-control studies are rising in frequency. The category 'other epidemiologic studies' used in this survey was very broad and included studies of all types, even descriptive reports. Thus, case-control studies are the predominant type of epidemiologic research and a major type of medical research; it is important that every effort be made to refine them.

III. STRENGTHS AND LIMITATIONS

It may be useful before suggesting improvements for case-control studies, to review their strengths and limitations as now perceived.

A major advantage claimed is that case-control studies can be done rapidly and inexpensively. This is not always true. Some case-control studies go on for several years and may cost several hundred thousand dollars. Moreover, non-concurrent follow-up studies are not very different from case-control studies in terms of cost and duration. It is true, then, to say that case-control studies are quick and inexpensive only in comparison with concurrent follow-up studies.

A second advantage suggested is that case-control studies are uniquely suited to the study of rare diseases; the rarer the disease the greater the relative advantage of the case-control approach. This is so, but a disease which is rare in a general population may not be rare in a special exposure group. In that circumstance the non-concurrent follow-up study again deserves consideration.

A third advantage is that a case-control study allows the evaluation of several different etiologic factors both as independent and interacting causes. This usually can not be done nearly as well in follow-up studies, whether concurrent or non-concurrent, and is a great strength of case-control studies.

Turning to limitations, it is often stated that case-control studies are not suitable for the study of rare exposures. However, if it is suspected that a rare exposure is a cause of a high proportion of a particular disease, then a case-control study is suitable. The case-control study of vaginal cancer in young women illustrates this as it clearly incriminated *in utero* exposure to diethylstilbestrol, a rather uncommon exposure [11].

A second limitation is that case-control studies allow estimation of relative rates but not of absolute rates. Again, qualification is needed. Numerous case-control studies have included incidence or prevalence surveys and have provided risk factor-specific absolute rates. Further, even when a survey is not included it may be possible to estimate the absolute rate of disease in the population studied and to infer the risk factor-specific absolute rates. This was done, e.g., in a study of oral contraceptives and thromboembolic and gall bladder disease conducted by the Boston Collaborative Drug Surveillance Program [12].

A third limitation of case-control studies is that they are highly susceptible to bias. This is true, and bias, especially selection bias, is probably the most serious potential problem in case-control studies. It is discussed at length below. Somewhat in compensation, the larger number of pertinent observations in a case-control than in a follow-up study makes the former less likely to be in error by chance.

IV. LESSER PROBLEMS

I now turn to areas where I believe improvements can be made to the betterment of the case-control study. First, some 'lesser problems'; those which are partially solved or where the need is more for refinement and promulgation of knowledge than for new knowledge.

Terminology

A first 'lesser problem' is the jungle of terminology in which we live. The epidemiology of chronic diseases is a new, vigorous science with its roots in many other sciences. Neologisms, inexactitudes, redundancies and superfluities of terminology are all to be expected and all exist. For example, the case-control study has at least three aliases:

a trohoc study, a case-referent study and a retrospective study, and the last of these names is in common use. But I see no need to depart from the term 'case-control' which was advocated in the early 1960's by Sartwell [13]. While not an overwhelming problem, our lack of a uniform terminology causes misunderstandings and makes the student's life difficult where it need not be. As the Lilienfelds have suggested, a glossary of epidemiologic terms should be developed [14]. And, this task should not be taken lightly, for to do a poor job might be worse than to do nothing.

Quantitative methods

As a second 'lesser problem' consider proposed needs for new quantitative methods for the analysis of case-control studies. I would suggest that in this area we have passed the point of diminishing returns. Techniques available both for multi-variate analyses and for stratified analysis are effective for describing association, for evaluating interaction and for evaluating and controlling confounding. What is needed is a major increase in the understanding, application and teaching of the tools we already have. To these ends, Day and Breslow will soon publish a monograph providing a comprehensive approach to the analysis of case-control studies based on the multiple logistic model. And Rothman and Boice will publish a book illustrating the use of a programmable desk-top calculator for most analyses useful to an epidemiologist.

Description of results

As a third problem consider the way in which the results of a case-control study are described. Mantel and Haenszel wrote "a primary goal is to reach the same conclusion in a (case-control) study as would have been obtained from a (concurrent follow-up) study" [4]. This could be improved by stating that the goal of a case-control study is to reach the correct conclusion. There is no need to invoke the results of a follow-up study, or of an experiment for that matter, as a benchmark. Nonetheless, the implied comparison of case-control study results with those from a follow-up study is useful. It focusses attention on the need to describe findings in terms which have biologic meaning. The provision only of exposure frequencies and related p-values is an inadequate, though still commonplace, endpoint for a case-control study. Full epidemiologic analysis assesses bias, confounding, causation and chance. Of these, chance is least important but still receives most attention. This has a historical basis but fortunately is now changing. The assessment of chance is receiving less attention and the extremely limited utility of the p-value for describing data is gradually coming to be recognized. How often must we remind ourselves that no p-value, however large, means that chance is an explanation of results? Similarly, no p-value, however small, excludes chance. Large p-value or small, the investigator is obliged to assess the extent to which confounding, bias and causality may explain his results. Specific objections to the p-value are these: (1) its use and the accompanying language often imply, erroneously, that causation can be established or refuted. Thus, one speaks of 'rejecting' or 'accepting' a hypothesis when a hypothesis can only have its credibility favorably or unfavorably modified; (2) the p-value is not informative as to the probable role of chance in a study with a negative, i.e. null, result. Indeed, there is little meaning that can be attached to a p-value related to a negative, or nearly negative result; (3) the p-value suppresses information because it is too complex a measure. It simultaneously reflects both a study's size and the observed strength of association and so does neither very well; (4) it is too readily misinterpreted. One still frequently sees non-statistically significant positive studies described as negative even by the investigator himself. And, this is a common error in review articles.

But, over-riding the objections to the p-value is a positive item which gives it the *coup de grâce*: there is a very superior alternative available. This is the confidence interval around the point estimate of effect. There used to be one reasonable objection to confidence limits, namely that they are tedious to compute in some situations encoun-

tered in case-control studies. But this is no longer true thanks to Miettinen's test-based procedure for estimating the limits [15].

In a related vein, I make two suggestions for improving the presentation of results of case-control studies. First, if p is used, provide a point estimate rather than describe it as 'less than 0.05'. This is inherently more informative and will also permit the reader who wishes to do so to 'back-calculate' the value of the significance test and then to estimate confidence limits. Second, whether p-values or confidence limits, or both, are used, consider using the one-tailed, instead of the two-tailed, statistic. In many epidemiologic studies the one-tailed statistic seems appropriate for there is usually interest in hypotheses which, before the study was done, were highly credible or related to a uni-directional alternative to the null state.

Interpretation

The final two 'lesser problems' are in the area of interpretation. The case-control study is still accused of being uniquely deficient when it comes to establishing causality. It is often implied, sometimes made explicit, that other kinds of studies, especially experiments, can establish causality but that case-control studies can not. For example, "the trouble with (a case-control study) is that it can never prove cause; it can show only an association..." [16]. True enough, but what kind of study can do more? The quotation implies that there are logical flaws or insuperable practical limitations in case-control studies which preclude a causal inference; it also implies that other studies can establish causation. Both implications are wrong, as an individual human being's evaluation of whether or not an association is causal is a complex judgmental affair in which the results of a good study, of whatever type, count heavily.

The second problem in interpretation returns us to the question of 'accepting' or 'rejecting' the null hypothesis. The null hypothesis serves a crucial role in statistics in providing the conceptual underpinnings for the development of significance tests and related procedures. It is important in epidemiology too. But there, its use should be confined to the evaluation of chance effects in producing results. Scientists, including epidemiologists, do not need the null hypothesis when it comes to *interpreting* results. This is because, as scientists, our purpose is only to describe nature. We have no need to accept or reject any hypothesis. The need to make such 'decisions' in order to plan a course of action falls to legislators, to regulators and to other policy makers. Scientists must assist in this but the decision-making process, which I do not denigrate, should be kept distinct from the process of generating the scientific bases for decision making. When a scientist does participate in decision-making it should be as a consultant on objective matters, not as a collaborator on subjective ones.

V. GREATER PROBLEMS

Case definition

Once a problem has been defined and a case-control study decided upon, attention usually moves to designating the cases. Considerable thought is given to practical matters relating to the criteria for a 'case' and to sources of such people. But what is the goal in attempting to define 'caseness'? I suggest that the goal should be to define a group of individuals who have a disease which is, insofar as possible, a homogeneous etiologic entity. Obviously it will be easier to perceive one causal web at a time rather than several. This is true whether the webs consist of one or more necessary factors. For example, it would be futile to study the epidemiology of 'cancer of the uterus'. But, if a distinction is made between adenocarcinoma of the uterine corpus and squamous cell carcinoma of the cervix and if research is directed to one or the other, progress can be made. We should go further in making such distinctions. We should not limit ourselves to defining diseases solely in terms of manifestational characteristics, no matter how many these are or how subtle we consider them to be. Such definitions may have sufficed in a bygone era in which there was, for practical purposes, a one-cause–one-manifestational-entity relationship. But they do not suffice for the 'dis-

eases' we face today. At any point in time we should use all existing knowledge, manifestational *and* epidemiologic, to help define the most homogeneous disease condition possible. Yet, in many studies little thought is given to this crucial factor. I shall not dwell on it here but this proposed severe restriction of the range of characteristics of cases included in a case-control study should have other, practical benefits; it should help reduce confounding and, since controls would also be so restricted, it should help reduce selection bias.

The suggestion to study a restricted group of cases and controls may seem to violate a 'tenet' of epidemiology, namely, that cases and controls should be representative of all those in a population. However, that is not so. I have been speaking only of defining a disease. The question of representativeness, or complete case ascertainment, is germane only within the context of some particular disease entity. Nonetheless, this issue of representativeness warrants discussion.

Representativeness

The ill-advised pursuit of representativeness causes unnecessary work and reduces the precision of epidemiologic studies. Further, accusations of non-representativeness may cast unjustified aspersions on good research. It has been considered that the pursuit of representativeness comes down to sacrificing precision to attain generalizability. If one takes a broad cross-section of cases one often finds that the value of the effect parameter varies over the range of some variable, say age. For example, the relative incidence (RI) of bladder cancer among smokers ranges from 1.5 for elderly men to 2.5 for young men [17]. A single study, of some specified size, of bladder cancer might give rise to a single estimate of the RI for men of all ages. This estimate, say 2.0, would probably apply to men of one age group or another but it would not apply to men in general. Moreover, the confidence limits around the estimate would be relatively wide. Alternatively, the study might provide four or five estimates, for four or five age groups, but each of these would be very imprecise. As another alternative, the study could have been restricted in the first place to men of one or two age groups and then a precise estimate of the RI, at least for those age groups, would have been obtained. As in this example, when analyzed, a representative study breaks down into a series of small studies from each of which an imprecise estimate of association is obtained. And, implicitly, when an estimate is seen to be imprecise, its perceived validity is reduced. It thus appeared that the proponents of representativeness were advocating the pursuit of widely generalizable study results irrespective of precision and validity. And, on this basis, their position has been attacked [18, 19]. But a reconsideration of the pro-representativeness position gives a different impression of the objective. It appears that advocates of representativeness [20] urge it *not* for reasons of generalizability at all. They urge it for validity. Specifically, they see the attainment of representativeness as one way of reducing selection bias, the crucial problem in a case-control study. They do not wish to see a case-control study based on a subgroup of cases which for some reason, unintended and unknown to the investigator, has had undue opportunity to sustain the exposure of interest. If this is correct, we should stop criticizing the pursuit of representativeness on the grounds that it is directed towards the wrong goal; it has an appropriate goal. If it is to be criticized, it is on several other bases; e.g. that it *de facto* leads to studies with imprecise results. I suggest that the problem of representativeness can be resolved by the study of groups of cases which are highly restricted in their characteristics, by definition, rather than by source of identification. If case groups of this sort are used one should be able to obtain a highly precise estimate of an association based on a representative series of cases of a particular type.

Control selection

Here, the problem of developing a control series is divided into four components: the number of control groups, the size of the group(s), the definition of the group(s) and the selection of individual subjects.

The question of how many control groups should be included in a case-control study is an area both for clarification of principles and for some empiric work. The reason is that at present we have two recommendations which are based on apparently reasonable principles but which are contradictory to one another. One recommendation is that, usually, a case-control study should have one control group, that group which is, in principle, best suited to the needs of the particular study. A second group should be added only if the first group has some specific known or suspect deficiency which can be offset by the second group. The alternative recommendation is that every case-control study should have at least two control groups. Then, if the results are similar when either group is used the validity of the study is enhanced. Resolution of this controversy would be very valuable. However, it should be recognized that if the use of two different control groups gives differing results it does not mean that efforts were wasted. The explanation of the discrepancy, if it can be deduced, may be very informative. For example, in a recent case-control study of Hodgkin's disease (HD) and tonsillectomy, the relative incidence was found to be 3.1 when the spouses of the cases were used as controls and 1.4 when the siblings were used [21]. These findings suggest that some correlate of the risk of having a tonsillectomy in childhood, which is over controlled-for by the use of sibling controls, is a cause of HD. Thus, the hypothesis emerges that some aspect of life style in childhood, perhaps exposure to infectious agents or some correlate of the frequent use of medical services, is a cause of HD [22].

The second question relates to the size of the control group. When the number of cases and controls available for a study is large and when the cost of gathering information from a case and a control is about equal then the selection ratio of controls to cases would be unity. The standard issues would then be invoked to develop an estimate of an acceptable minimum study size. The question becomes more complex when, for whatever reason, the size of either group is severely limited or the cost of obtaining information is greater for one type of subject than for the other. For example, it occurs frequently that the number of cases available is fixed at a relatively small number. In such a circumstance the selection ratio should be increased so that there are two, three or even four controls per case. This is obvious but it is not commonly done. It is distressing still to see otherwise good case-control studies which are non-persuasive because of their unnecessarily small size. The selection ratio should be permitted to vary according to the circumstances of each study. But, one must be wary; it is wise to stay within the bounds of 4:1, perhaps 5:1, except when the data are 'free'. The reasons for this have been presented by Gail *et al.* [23] and by Walter [24]. Most of the justification is based on the small increase in statistical power as the ratio increases beyond four. It is worth noting that if more than one control group is used, not all groups need be of the same size.

The third issue is the unique and the truly large problem of the case-control study, the selection of the control group. This is the issue of avoiding selection bias. It is the problem of assuring that under the null state cases and controls would have been equally exposed to the factor of interest. The question of selection bias can not be entertained with respect to the case or the control series; it is the question of their comparability. But since one usually chooses the case series first the issue of avoiding selection bias is the question of choosing an appropriate control group. With respect to this general topic I shall address three relatively specific issues.

The first issue concerns a suggestion often made to reduce selection bias, namely that the controls should undergo the same diagnostic procedure as the cases. This is intended to overcome selection of cases who are excessive users of medical services and the index of suspicion which is a result of the physician's knowledge of the patient's exposure history. The suggestion translates into two seemingly similar, but in fact entirely different, courses of action. One course of action, to which I see no theoretical objection, is to select controls, however one will, and then to subject them to the diagnostic procedure. This is expensive, poses practical difficulties and for some procedures

would be ethically unacceptable. This might eliminate perhaps as much as 5% of controls who are, in fact, cases-to-be. It would be to little avail and I know of no case-control study in which this procedure was followed. The second course of action has been taken [25]. This is to choose controls from among people who have already undergone the same diagnostic procedure as the cases but who were found to have no disease or a disease different from that of the cases. This is an inappropriate control group because agents which cause one disease in an organ often, perhaps usually, cause other diseases of that organ. Do this for lung cancer and persons with chronic bronchitis will be the controls. You will still perceive an association of lung cancer with smoking but it will be muted because smoking causes bronchitis. Despite this difficulty, the use of a diagnostic register as a source of controls may prove to be a valuable way to control the possible 'medical consumerism' selection bias described above. However, to be appropriate, such rosters of potential controls should relate to procedures for the diagnosis of conditions of organs other than that which is the site of the disease which afflicts the cases.

A second issue regarding control definition relates to the recurrent error that 'the controls must be like the cases in every respect other than having the disease of interest'. This misconception springs anew in the mind of every student and it appears in recent text books. Its historical basis is clear; it comes from the axiom of experimental research that the control subjects must be treated in every respect like the exposed subjects. But in a case-control study the old axiom is inapplicable. The consequence of selecting the controls to be like the cases with respect to some correlate of the exposure under study, but which correlate is not itself a risk factor, that is 'overmatching', is now well recognized [26]. This leads to an imprecise estimate of the measure of effect and, unless an appropriate analysis is done, the estimate will also be biased towards the null value. Overmatching usually also entails considerable increase in the duration and cost of the study.

A third issue regarding the definition of controls and a major factor in case-control studies is the source of the control group. Most studies use either hospital patients or the general population as the source of controls. Much less often used are restricted population groups, e.g. neighbors of cases or special groups such as associates or relatives of cases.

The general population has a major strength as a control group. Such controls will be especially comparable to the cases when a population-based series of cases has been assembled. In many ways, this often makes for the most persuasive type of case-control study. This is because of the high comparability of the two series and because a rather high level of generalizability of results will be achieved. However, there are two serious disadvantages associated with using the general population as a control group. For one, it can be extremely expensive and time consuming to select such a group. For another, the individuals selected often are not cooperative and response tends to be poorer than that of other types of controls. This second disadvantage is especially important because it detracts from the presumed major strength of a general population control group.

The use of hospital patients as a control group has several advantages. Such people are usually readily available, have time to spare and are cooperative. Moreover, since they are hospitalized (or recently have been) they may have a 'mental set' similar to that of the cases. This should reduce the problem of selective recall of events, one of the most serious potential problems in a case-control study. The use of hospital patients as controls also may make the cases and controls similar with respect to the determinants of hospitalization. This is probably useful if the cases have a disease for which hospitalization is elective. The use of hospital patients as controls has one possibly serious limitation. The controls may be in a hospital for a condition which shares etiologic features with that of the disease under study. To minimize this problem, controls should be selected from many diagnostic categories.

There is no ready way to select one control group over another. Such selection

depends upon an understanding of the factors under study and the way they relate to the characteristics of the groups under consideration.

The fourth and last major concern in the selection of controls pertains to the selection of the individual subjects from among all those who are eligible. One very important question which arises relates to the extent to which the controls should be matched to the cases on an individual basis. The approaches to answering this question will be presented by others in this Symposium. A second question which always comes up in relation to a hospital-based control series pertains to who may be excluded from the control series. Should we exclude no such persons? Or, should we exclude persons who have conditions known or suspected to be related to the factors suspected to cause the disease of the cases? If so, what if the study is exploratory and relates to innumerable possible causes, many of which are not highly credible? If exclusions are permitted, would one exclude a potential control on the basis of his current condition or total medical history? If the latter, would one also exclude cases who have previously had such conditions? Here is an area where some guidelines are needed.

Data collection

A fourth and very important 'greater problem', that of information gathering with its opportunities for observer and subject bias, will not be discussed. I just mention that in studies which rely on anamnestic information the possibility of selective recall, usually forgetfulness on the part of controls and/or exaggerated remembrance on the part of cases, is one of the most serious and formidable deficiencies in case-control studies. Methods to reduce this bias, such as re-interview and validity checks, are useful even if not entirely satisfactory. Unfortunately, in some instances, even these buttresses are unavailable.

Data analysis

In this area I would like to point out three problems. First, if one has used more than one control group, what criteria should be met before those groups might usefully be merged? Can one take a formal statistical approach to this problem? I believe this could be worked out but that it has not been done yet. Certainly, the idea that one would merge two control groups if their exposure frequencies are not 'significantly' different from one another but not to do so if they are, is quite unappealing. Less quantitative, but more appealing, is the notion that the merger would occur, or not, depending on the investigator's subjective assessment of the distortion introduced by the use of an effect measure based on the two series combined as compared to the two single-group measures. While it is easy to say that merger should never occur, this raises a practical problem when the two groups give the same result. It is wasteful to discard one group but tedious, in a report, to make reference repeatedly to two sets of essentially identical results. In any event, a defensible action is to not merge the data from two control groups but only the inferences that one makes from the results of the two case-control comparisons.

A second problem in analysis relates to the use of 'heterogeneity' testing. Mantel *et al.* have recently reviewed this problem and wisely suggest that heterogeneity testing be used with extreme caution [27]. The problems of scaling to which Mantel refers are but the beginnings of a labyrinth of logical difficulties which makes the interpretation of the results of a heterogeneity test treacherous, particularly when the effect measure evaluated is a series of relative incidences. For now, the judgment as to whether or not two or more factors are interacting or modifying one another's effect seems better left largely a subjective matter.

A third issue in the analysis of case-control studies relates to the so-called 'multiple-comparison' problem. Consider first the circumstance where there are more than two study groups (as when several control groups are used) and many inter-comparisons are possible, but only one variable is at issue. In this instance the nominal p-values which one obtains as a result of these comparisons are too low. That is, they should

be adjusted upwards because several non-independent comparisons have been made. Consider now, however, the circumstance where there are only two study groups but many variables. Somehow, I think by an analogy with the first circumstance, some persons suggest that the *p*-values resulting from these many comparisons are not valid and should also be adjusted upwards. The fact that many comparisons have been made and, thus, that some may be expected to be significant by chance alone, is supposed to detract from each of the *p*-values obtained. It is the same as saying that an association is penalized because it emerged in a large, rather than in a small, study. This is bothersome because, under a null state, the *p*-value has a 5% chance of taking on the value of 0.05 by chance alone whether it relates to the only variable evaluated in a study or to one of hundreds. A valid reason for not attaching great importance to a finding which emerges in a 'fishing expedition' is not that it relates to one of many variables but rather that, usually, the association in question had little advance credibility. In every study, every association should be evaluated on its own merits; its prior credibility and its features in the study at hand. The number of other variables is irrelevant.

Most of the problems to which I have alluded, relate to defining and avoiding 'bias' in case-control studies. Probably, some useful general guidelines can be developed. But there are biases peculiar to each type of disease, to each type of exposure and even to every particular study. In addition to forming general guidelines, perhaps we can attack an intermediate area of defining biases peculiar to certain types of studies. For example, Jick and Vessey have done this for case-control studies of drug exposures [19].

VII. SPECIFIC RECOMMENDATIONS

These recommendations have been tabulated in an outline form which, in general, follows the chronology of a case-control study:

(A) *General considerations*

(1) Develop a sanctioned terminology.
(2) Improve education regarding existing epidemiologic methods.
(3) Develop a catalog of biases, their sources and methods of control for specific study types.

(B) *Designing a study*

(1) Cases—improve methods for defining a single etiologic entity.
(2) Controls—clarify criteria for inclusion of more than one group;
—clarify criteria for excluding specific individuals;
—do not use people with disease of same organ as the disease of interest;
—choose selection ratio best suited to particular study.

(C) *Conducting a study*

(1) Develop methods to reduce selective recall.
(2) Develop methods to conceal sources of data from study staff.

(D) *Data analysis*

(1) Establish criteria for merger of control groups.
(2) Use measure of effect, not exposure frequencies.
(3) Use confidence limits, not *p*-values.
(4) Use one-sided statistics more often.
(5) Rarely do heterogeneity testing.

(E) *Interpretations*

(1) Do not denigrate value of a finding because it occurs in a study of many variables.
(2) Describe nature, do not 'make decisions'.

VIII. PROSPECTS

The use and value of case-control studies will increase in the years ahead. One stimulus to this will be the finding of new objectives for it. Up to now the case-control study has been vitually restricted to etiologic investigations. But this is unnecessary. For example, the case-control study could be used to evaluate preventive medical services and this has been attempted for the pap smear.

Besides new objectives, new approaches will give strength to the case-control study. Several interesting approaches are being developed. For example, we now occasionally see the multi-disease study, a simultaneous study of cases of several different types of disease. In the past when these were done the various case series usually used one another as controls, but newer studies will have true controls. We also now occasionally see the very efficient case-control-within-a-cohort study. This should greatly reduce both selection bias and selective recall. It is also extremely efficient from the point of view of information gained per dollar spent [28].

The value of the case-control study will also increase because our understanding of its fundamental nature is increasing. It was only 3 yr ago that Miettinen clarified the nature of the parameters estimated in a case-control study and also showed the usual irrelevance of the 'rare-disease' assumption [15]. If progress is still being made in such fundamental areas we can only presume that the case-control technique will improve in the years ahead. To put it another way, if Mantel and Haenszel recognized 20 yr ago that a case-control study could give the same result as a follow-up study, then we should be far-sighted enough to see today that the case-control study can, in many situations, replace the follow-up study.

REFERENCES

1. Cornfield J: A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast and cervix. **J Natl Can Inst** 11: 1269–1275, 1951
2. Dorn HF: Some applications of biometry in the collection and evaluation of medical data. **J Chron Dis** 1: 638–669, 1955
3. Dorn HF: Some problems arising in prospective and retrospective studies of the etiology of disease. **N Engl J Med** 261: 571–579, 1959
4. Mantel N, Haenszel W: Statistical aspects of data from retrospective studies of disease. **J Natl Can Inst** 22: 719–748, 1959
5. Cornfield J, Haenszel W: Some aspects of retrospective studies. **J Chron Dis** 11: 523–534, 1960
6. Anderson G: Personal communication
7. Lane-Claypon JE: A further report on cancer of the breast. Reports on Public Health and Medical Subjects 32: Min of Hlth, London: H.M.S.O., 1926
8. Doll R, Hill AB: A study of the aetiology of carcinoma of the lung. **Brit Med J** 2: 1271–1286, 1952
9. Wynder EL, Graham EA: Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. A study of six hundred and eighty-four proved cases. **JAMA** 143: 329–336, 1950
10. Bailar JC III, Anthony GP: Most cited papers of the Journal of the National Cancer Institute 1962–75. **J Natl Can Inst** 59: 709–714, 1977
11. Herbst AL, Ulfelder H, Poskanzer DC: Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. **N Engl J Med** 284: 878–881, 1971
12. Boston Collaborative Drug Surveillance Program: Oral contraceptives and venous thromboembolic disease, surgically confirmed gall-bladder disease, and breast tumours. **The Lancet** 1: 1399–1404, 1973
13. Sartwell PE: Retrospective studies: A review for the clinician. **Ann Intern Med** 81: 381–386, 1974
14. Lilienfeld DE, Lilienfeld AM: Re: 'Causes'. (Letter to the Editor). **Amer J Epid** 106: 84–85, 1977
15. Miettinen O: Estimability and estimation in case-referent studies. **Amer J Epid** 103: 226–235, 1976
16. Gordon GS, Greenberg BG: Exogenous estrogens and endometrial cancer. **Postgrad Med** 59: 66–77, 1976
17. Cole P, Monson RR, Haning H, Friedell GH: Smoking and cancer of the lower urinary tract. **N Engl J Med** 284: 129–134, 1971
18. Editorial: Oestrogen therapy and endometrial cancer. **Brit Med J** 6081: 209–210, 1977
19. Jick H, Vessey MP: Case-control studies in the evaluation of drug-induced illness. **Amer J Epid** 107: 1–7, 1978
20. Feinstein AR: Clinical Biostatistics, XX: The epidemiologic trohoc, the ablative risk ratio, and 'retrospective' research. **Clin Pharmacol Ther** 14: 291–307, 1973
21. Gutensohn N, Li F, Johnson R, Cole P: Hodgkin's disease, tonsillectomy and family size. **N Engl J Med** 292: 22–25, 1975
22. Gutensohn N, Cole P: The epidemiology of Hodgkin's disease among the young. **Int J Can** 19: 595–604, 1977
23. Gail M, Williams R, Byar DP, Brown C: How many controls? **J Chron Dis** 29: 723–731, 1976

24. Walter SD: Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. **Amer J Epid** 105: 387–397, 1977
25. Horwitz RI, Feinstein AR: New methods of sampling and analysis to remove bias in case-control research: no association found for estrogens and endometrial cancer (abstracts of papers presented at the Tenth Annual Meeting of the Society for Epidemiologic Research). **Amer J Epid** 106: 251, 1977
26. Miettinen OS: Matching and design efficiency in retrospective studies. **Amer J Epid** 91: 111–118, 1970
27. Mantel N, Brown C, Byar DP: Tests for homogeneity of effect in an epidemiologic investigation. **Amer J Epid** 106: 125–129, 1977
28. Liddell FDK, McDonald JC, Thomas DC: Methods of cohort analysis: Appraisal by application to asbestos mining. **J Roy Stat Soc, Ser A** 140: 469–491, 1977