J Chron Dis Vol. 32, pp. 51 to 63
Pergamon Press Ltd 1979. Printed in Great Britain

BIAS IN ANALYTIC RESEARCH

DAVID L. SACKETT

INTRODUCTION

CASE-CONTROL studies are highly attractive. They can be executed quickly and at low cost, even when the disorders of interest are rare. Furthermore, the execution of pilot case-control studies is becoming automated; strategies have been devised for the 'computer scanning' of large files of hospital admission diagnoses and prior drug exposures, with more detailed analyses carried out in the same data set on an *ad hoc* basis [1]. As evidence of their growing popularity, when one original article was randomly selected from each issue of **The New England Journal of Medicine, The Lancet,** and the **Journal of the American Medical Association** for the years, 1956, 1966 and 1976, the proportion reporting case-control analytic studies increased fourfold over these two decades (2–8%) whereas the proportion reporting cohort analytic studies fell by half (30–15%); incidentally, a general trend toward fewer study subjects but more study authors was also noted [2].

If an ebullition of case-control studies is in progress, a review of their merits and shortcomings is of more than academic interest, and this symposium was well-timed. Because this meeting also coincided with the completion of some work we had been doing on biases in analytic research (Appendix 3), I offered to summarize a portion of this work for presentation and discussion here.

A first draft of a catalog of biases which may distort the design, execution, analysis and interpretation of research appears as an appendix to this paper (additions, corrections and citations of examples would be welcomed by the author). For this paper, I have considered those biases which arise in analytic studies and have focused on two subsets which affect the specification and selection of the study sample and the measurement of exposures and outcomes, since these attributes most clearly distinguish the case-control study from its relatives.* Furthermore, I have included occasional discussions of cohort analytic studies because they represent a common, alternative, subexperimental approach to determining causation. Finally, after describing the prospects for the prevention (or at least the measurement) of these biases in these two forms of analytic studies, this paper closes with suggestions for further methodologic research.

DEFINITIONS AND EXAMPLES

To date we† have cataloged 35 biases that arise in sampling and measurement (see the appendix) and nine of these will be discussed here:

(1) *Prevalence-incidence (Neyman) bias*

A late look at those exposed (or affected) early will miss fatal and other short episodes,

*In discussing the biases of sampling I have tried to avoid introducing the biases of rhetoric; the latter, though 'good theater', are both inappropriate for this symposium and better discussed elsewhere [4].

†The catalog was initiated by a clinical epidemiology graduate student, JoAnne Chiavetta; it was benefitted from the contributions of a number of colleagues (especially John C. Sinclair) and other publications (especially references [5] and [6]).

TABLE 1. COHORT VS CASE-CONTROL ESTIMATES OF THE RELATIVE ODDS OF CORONARY HEART DISEASE AMONG FRAMINGHAM MEN WITH AND WITHOUT HYPERCHOLESTEROLEMIA (UPPER QUARTILE OF THE DISTRIBUTION OF SERUM CHOLESTEROL)

	Cohort study			Case-control study		
	Developed coronary heart disease by exam 6	Did not develop coronary heart disease by exam 6	Totals	Coronary heart disease present at exam 6	Free of coronary heart disease at exam 6	Totals
Highest quartile of serum cholesterol*	85	462	547	38	34	72
Lower 3 quartiles of serum cholesterol*	116	1511	1627	113	117	230
Totals	201	1973	2174	151	151	302
Relative odds (cross-products): 2.40				Relative odds (cross-products): 1.16		

*Cholesterol as measured at Exam 1 in the cohort study and Exam 6 in the case-control study. Table derived from data in Friedman *et al.* [8].

plus mild or silent cases and cases in which evidence of exposure disappears with disease onset [7].

A disorder which illustrates the properties of the prevalence-incidence bias is clinical coronary heart disease. We recognize that the high case-fatality rate in the early moments of clinically-manifest myocardial infarction may invalidate the study of possible etiologic factors among even short-term survivors. Similarly, we acknowledge the existence of the 'silent' myocardial infarction, as well as the potential for all clinical and paraclinical indexes of myocardial cell death (including the electrocardiogram) to return to normal after the event. Finally, we recognize that evidence of coronary risk may disappear with disease onset. This is commonly seen clinically in hypertensive patients, and was demonstrated for hypercholesterolemia by Friedman *et al.* in the Framingham Study, as shown in Table 1 [8].

The latter investigators found a similar pattern of change in relative odds when the cohort component was restricted to those who survived to Exam 6 (ruling out selective mortality as the cause), and suggested that coronary patients might be 'more careful about their diet' after the onset of clinically manifest disease.

The prevalence-incidence bias is of at least potential importance in any analytic investigation where a time gap exists between exposure and the selection of study subjects. Moreover, this bias may distort relative odds in either direction. In the foregoing examples, its effect was a spurious decrease in relative odds. If, on the other hand, an exposure led to selective survival (rather than selective mortality), the relative odds calculated from a later case-control study would be spuriously raised.

(2) *Admission rate (Berkson) bias*

If the admission rates of exposed and unexposed cases and controls differ, their relative odds of exposure to the putative cause will be distorted in hospital-based studies [9].

Berkson's 'paradox' (for this is the term preferred by its author) is, in itself, a paradox. Although it was described over 30 yr ago and has been cited in a great number of papers since that time, it was not empirically demonstrated until quite recently [10, 3]; thus, in a 1974 review of case-control studies, Sartwell observed that 'its practical importance has not been established' [11].

We have recently tested for Berkson's bias in a body of household interviews performed upon random samples of the general population [10, 3]. Because these interviews included information both about diseases and about recent hospitalizations, it was possible to calculate the relative odds of several diseases (given specific prior exposures)

TABLE 2. THE RELATIVE ODDS OF DISEASE OF THE BONES AND ORGANS OF MOVEMENT WITH AND WITHOUT RESPIRATORY DISEASE*

		In the general population			In the subset who were in the hospital in the prior 6 months		
		Disease of bones and organs of movement			Disease of bones and organs of movement		
		Yes	No	Totals	Yes	No	Totals
Respiratory disease	Yes	17	207	224	5	15	20
	No	184	2376	2560	18	219	237
	Totals	201	2583	2784	23	234	257
		Relative odds (cross-products): 1.06			Relative odds (cross-products): 4.06		

*Adapted from Roberts *et al.* [3].

both in the general population and in that subset of the same general population who had been hospitalized in the previous 6 months. Two examples of these analyses appear in Tables 2 and 3.

As seen in these examples, relative odds may be spuriously increased or reduced by the admission rate bias, and comparisons of the upper left-hand cells between the halves of Tables 2 and 3 demonstrates that individuals with both conditions may have either relatively high (Table 2) or low (Table 3) admission rates.

The admission rate bias may have many causes (the burden of symptoms, access to care, popularity of disorders and institutions, etc.) and is of at least potential importance in any hospital- or practice-based study of etiology. Because it is precisely these settings that makes possible the study of diseases that are rare or late, this bias is central to the execution of case-control studies.

(3) *Unmasking (detection signal) bias*

An innocent exposure may become suspect if, rather than causing a disease, it causes a sign or symptom which precipitates a search for the disease.

Increasing attention to this bias has occurred in the course of considering the relation between post-menopausal estrogens and endometrial cancer. Horwitz and Feinstein considered the possibility that estrogens might cause the *search* for endometrial cancer (by causing symptomless patients to bleed) rather than the cancer itself, and compared the relative odds obtained from patients in a tumor registry (51% of whom presented with bleeding; the left-hand panel of Table 4) with that obtained from a registry of patients who had undergone dilatation and curretage or hysterectomy at the same institution (76% whom had presented with uterine bleeding; the right-hand panel of Table 4) [12].

TABLE 3. THE RELATIVE ODDS OF FATIGUE WITH AND WITHOUT PRIOR ALLERGIC OR METABOLIC DISEASE*

		In the general population			In the subset who were in the hospital in the prior 6 months		
		Fatigue			Fatigue		
		Yes	No	Totals	Yes	No	Totals
Allergic and metabolic disease	Yes	13	136	149	1	21	22
	No	127	2508	2635	27	208	235
	Totals	140	2644	2784	28	229	257
		Relative odds (cross-products): 1.89			Relative odds (cross-products): 0.37		

*Adapted from Roberts *et al.* [3].

TABLE 4. RELATIVE ODDS OF ENDOMETRIAL CANCER WITH AND WITHOUT EXPOSURE TO POST-MENOPAUSAL ESTROGENS; TWO STUDIES AT THE SAME INSTITUTION*

	Tumor registry			D and C/Hysterectomy registry		
	Endometrial cancer			Endometrial cancer		
	Yes	No	Totals	Yes	No	Totals
Post-menopausal estrogens	45	7	52	59	42	101
No post-menopausal estrogens	72	110	182	89	106	195
Totals	117	117	234	148	148	296
	Relative odds (cross-products): 9.8			Relative odds (cross-products): 1.7		

*Adapted from Horwitz and Feinstein [12].

The results are consistent with the performance of the unmasking bias, and it should be noted that the proportion of cases exposed to estrogens was the same in both studies although, as expected, estrogen users were more likely to have Stage I cancer (79%) than cases who had not used estrogens (58%).

Thus, the unmasking bias may lead to spuriously increased estimates of relative odds. On the other hand, in attempting to prevent this bias could the restriction of cases and controls to only those patients who have undergone identical detection maneuvers (a standard approach in cohort analytic studies and experiments) lead to 'over-matching' in case-control studies [13]? This latter issue remains to be resolved.

(4) *Non-respondent bias*

Non-respondents (or 'late-comers') from a specified sample may exhibit exposures or outcomes which differ from those of respondents (or 'early comers'); the antithetical bias is called the 'volunteer' bias.

This bias is ubiquitous in descriptive, analytic and experimental research and has been demonstrated repeatedly among cigarette smokers. For example, in a mailed questionnaire study of the smoking habits of U.S. veterans, Seltzer *et al.* noted that 85% of non-smokers, but only 67% of cigarette smokers, returned the questionnaire within 30 days, with an intermediate return rate for pipe and cigar smokers [14].

The effect of the non-respondent bias upon relative odds is obvious and serves as the basis for repeated admonitions both to achieve response rates of at least 80% and to compare responders and non-responders.

(5) *Membership bias*

Membership in a group (the employed, joggers, etc.) may imply a degree of health which differs systematically from that of the general population.

TABLE 5. THE RELATIVE ODDS OF CORONARY DEATH WITH AND WITHOUT EMPLOYMENT IN A PHYSICALLY ACTIVE OCCUPATION; A CASE-CONTROL STUDY*

		Deaths due to coronary heart disease	Deaths from conditions not associated with coronary disease	
Physical activity characteristics of most recent occupation	Heavy	194	668	862
	Light or active	840	2029	2869
		1034	2697	3731
Relative odds (cross-products): 0.70				

*Adapted from Morris and Crawford [15].

TABLE 6. THE RELATIVE ODDS OF RECURRENT MYOCARDIAL INFARCTION WITH AND WITHOUT PARTICIPATION IN A GRADUATED EXERCISE PROGRAM FOLLOWING AN INITIAL MYOCARDIAL INFARCTION; A COHORT ANALYTIC STUDY*

		Recurrent myocardial infarction		
		Yes	No	
Participation in graduated exercise	Yes	7	59	66
	No	18	46	64
		25	105	130
Relative odds (cross-products): 0.30				

*Adapted from Rechnitzer *et al.* [16].

The most topical example of the membership bias (at least in North America) is the jogger. The hypothesis that vigorous physical activity protected against coronary heart disease received its initial support from case-control studies such as that shown in Table 5 [15].

When this hypothesis was further tested among cohorts of survivors of myocardial infarction who did and did not engage in graduated exercise, further support was gained as shown in Table 6 [16].

However, when these encouraging results from case-control and cohort analytic studies were tested in a randomized trial in which eligible survivors of myocardial infarction were randomly allocated to twice weekly endurance training or recreational activities which would not produce a 'training effect', as shown in Table 7, the value of physical activity could not be substantiated [17].

In addition to demonstrating the membership bias, these examples indicate that this bias may affect cohort as well as case-control analytic studies.

(6) *Diagnostic suspicion bias*

A knowledge of the subject's prior exposure to a putative cause (ethnicity, taking a certain drug, having a second disorder, being exposed in an epidemic) may influence both the intensity and the outcome of the diagnostic process.

A frequent caution to the clinician-in-training who is learning clinical skills, this bias has been explored only recently within an epidemiologic context. Fox and White, concerned that physicians who were aware of the putative causal relation between working in the rubber industry and bladder cancer might be influenced by this knowledge, tried to determine whether bladder cancer was understated for men working in other occupations [18]. This bias is usually associated with the cohort analytic study, but it may also affect the generation of cases and controls in a case-control study if the putative causal factor has received widespread publicity.

TABLE 7. THE RELATIVE ODDS OF RECURRENT MYOCARDIAL INFARCTION WITH AND WITHOUT ENDURANCE TRAINING IN A RANDOMIZED TRIAL*

		Recurrent myocardial infarction		
		Yes	No	
Randomly allocated to undergo endurance training	Yes	28	359	387
	No	21	345	366
		49	704	753
Relative odds (cross-products): 1.28				

*Adapted from Rechnitzer *et al.* [17].

TABLE 8. THE INFLUENCE OF THE INTENSITY OF SEARCHING FOR EXPOSURE UPON REPORTED RATES OF EXPOSURE*

	Prior exposure to irradiation	
	Upon routine questioning and records (%)	Upon intensive questioning and search (%)
36 cases of Nishiyama *et al.* [19]	28	47
22 cases of Raventos *et al.* [20]	0	50

*Adapted from Nishiyama *et al.* [19] and Raventos *et al.* [20].

(7) *Exposure suspicion bias*

A knowledge of the patient's disease status may influence both the intensity and outcome of a search for exposure to the putative cause.

Another bias well known to clinicians, the exposure suspicion bias may operate whenever patients appear with disorders whose 'causes' are 'known'. The magnitude of this bias was shown in studies of thyroid cancer among children in which, depending upon the intensity of the search for prior irradiation, markedly different rates of exposure were reported; this is shown in Table 8 [19, 20].

(8) *Recall bias*

Questions about specific exposures may be asked several times of cases but only once of controls.

The recall of cases and controls may differ both in amount and in accuracy. For example, in questioning mothers whose recent pregnancies had ended in fetal death or malformation (cases) and a matched group of mothers whose pregnancies ended normally (controls) it was found that 28% of the former, but only 20% of the latter, reported exposure to drugs which could not be substantiated either in earlier prospective interviews or in other health records [21].

The recall bias may be most marked when the exposure of interest is rare or when controls are drawn from the community rather than from hospitalized patients.

(9) *Family information bias*

The flow of family information about exposures and illnesses is stimulated by, and directed to, a new case in its midst.

The family information bias was demonstrated by Schull and Cobb in their study of whether rheumatoid arthritis clusters in families [22]. When these investigators asked

TABLE 9. FAMILY HISTORY OF ARTHRITIS AMONG INDIVIDUALS WITH AND WITHOUT RHEUMATOID ARTHRITIS*

	Person reporting family history	
	19 persons with rheumatoid arthritis (%)	201 persons free of rheumatoid arthritis (%)
%Reporting neither parent had arthritis	16	55
% Reporting one parent had arthritis	53	37
% Reporting both parents had arthritis	31	8
	100	100

*Adapted from Schull and Cobb [22].

TABLE 10. EFFECT OF THE SOURCE OF FAMILY INFORMATION UPON THE RESULTS OF THE FAMILY HISTORY*

	Sibling providing family history	
	Sibling with rheumatoid arthritis (%)	Sibling free of rheumatoid arthritis (%)
% Reporting neither parent had arthritis	27	50
% Reporting one parent had arthritis	58	42
% Reporting both parents had arthritis	15	8
	100	100

*Adapted from Schull and Cobb [22].

groups of individuals with and without rheumatoid arthritis whether their parents had arthritis, they obtained the results shown in Table 9 which suggested that the disorder did, indeed, 'run in families'.

However, when these investigators compared family histories on the *same parents*, obtained by independently asking 40 individuals with rheumatoid arthritis and their unaffected siblings whether their parents had arthritis, they obtained the remarkable results shown in Table 10 [22].

Thus, the family history (and, by analogy, other historical information) may vary markedly depending upon whether the individual providing the information is a case or a control, and the effect of this bias upon the relative odds may be profound.

In summary, nine biases of special importance in analytic studies have been drawn from a much larger number and have been described. Their effects upon the relative odds observed in case-control (and, for comparison, cohort) analytic studies are summarized in Table 11. The subsequent sections of this essay will consider their preventability, measurability, and impact upon the validity of case-control studies, plus some proposals for future methodologic research.

PREVENTION AND MEASUREMENT OF BIAS IN ANALYTIC STUDIES

In discussing bias in observational research before the Royal Statistical Society, Cochran summarized the general state of affairs:

TABLE 11. EFFECT OF NINE BIASES UPON RELATIVE ODDS OBSERVED IN CASE-CONTROL AND COHORT ANALYTIC STUDIES

Types of bias	Effect on relative odds	
	Case-control	Cohort
Sampling biases:		
Prevalence-incidence bias	↓ or ↑	(↓ or ↑)
Admission rate bias	↑ or ↓	DNA*
Unmasking bias	↑	↑
Non-respondent bias	↓ or ↑	↓ or ↑
Membership bias	↓ or ↑	↓ or ↑
Measurement biases:		
Diagnostic suspicion bias	(↑)	↑
Exposure suspicion bias	↑	(↓) or DNA
Recall bias	↑	DNA
Family information bias	↑	DNA

*DNA: Does Not Apply.
( ): unlikely to occur.

TABLE 12. THE PREVENTABILITY AND MEASURABILITY OF SELECTED BIASES IN CASE-CONTROL AND COHORT ANALYTIC STUDIES

	Case-control		Cohort	
	Preventable?	Measurable?	Preventable?	Measurable?
Sampling biases:				
Prevalence-incidence bias	No	Partially	Yes	Yes
Admission rate bias	Not without sacrificing its value	No	DNA*	DNA*
Unmasking bias	Yes, but over-match?	Yes	Yes	Yes
Non-respondent bias	Yes	Yes	Yes	Yes
Membership bias	No	Partially	No	Partially
Measurement biases:				
Diagnostic suspicion bias	DNA*	DNA*	Yes	Yes
Exposure suspicion bias	Yes	Yes	DNA*	DNA*
Recall bias	Yes	Yes	DNA*	DNA*
Family information bias	Yes	Yes	DNA*	DNA*

*DNA: Does Not Apply.

> '... neither the investigator nor the appraising committee can suggest a method reducing these biases (except that in some cases a completely different type of study might be less vulnerable to bias)... This type of proposal leaves the statistician frustrated...' [23]; to which Brown has added: 'Think what it does to the investigator!' [24].

I do not believe that the situation is as dim as this, and that several biases can be both prevented and measured. My estimates of their preventability and measurability are summarized in Table 12 (again, cohort analytic studies are included for comparison).

At least among those biases discussed in this paper, measurement biases are easier to prevent and measure than sampling biases. In the case of the former, effective strategies have included 'blinding' interviews to the subjects' diagnoses (or executing interviews about exposure prior to definitive diagnosis), establishing explicit, objective criteria for exposures and outcomes, and obtaining information about exposure from independent sources that are unaffected by memory or by the flow of family information.

However, sampling biases present a much more difficult problem. The non-respondent bias can be prevented by achieving high response rates ($\geqslant 80\%$ by convention), but it is in the other sampling biases that the case-control analytic study pays the price for its time-and-cost advantages.

As Berkson wrote in discussing the admission rate bias*: "there does not appear to be any ready way of correcting the spurious correlation existing in the hospital population by any device that does not involve the acquistion of data which would themselves answer the primary question" [9]. Thus, although one could prevent the admission rate bias by conducting an analytic survey in the general population, the result is no longer a case-control study, and loses its time-and-cost advantages.

The prevalence-incidence bias presents an analogous quandary. The exact composition of the groups of exposed and unexposed individuals from which cases and controls are sampled is not known in the case-control study. Thus, neither the comparability nor the attrition of the former can be known in this analytic design; alternatively, the strategy which overcomes this bias by identifying the comparability and attrition of these groups of exposed and unexposed individuals is no longer the case-control study, but the cohort analytic study.

The restriction of cases and controls to only those individuals who have undergone identical diagnostic examinations constitutes a useful preventive strategy borrowed from cohort analytic studies and experiments. However, Horwitz and Feinstein have suggested that the *clinical indications* for these diagnostic tests, as well as the tests themselves,

*Berkson's original paper was restricted to the consequences of combining probabilities, plus 'the burden of symptoms'. Our definition of the admission rate bias goes beyond these to consider all factors leading to differential admission rates. However, I believe that the quotation holds for both definitions.

should be identical in case-control studies [12], a restriction which raises the possibility of over-matching [13]. Perhaps this question can be resolved at this symposium.

Finally, prevention of the membership bias necessitates the recognition of, and the matching or adjustment for, all important confounding variables. Despite my respect for advances in matching and adjustment, I don't think we know enough about the determinants of membership to recognize them and effectively prevent this bias.

Both case-control and cohort analytic studies are susceptible to bias but, if the nine biases selected for detailed discussion here are appropriate to the issue, it appears that of the two designs the case-control strategy is both affected by more sources of bias and less able to defend against them.

If this assessment is valid, and in the absence of experimental evidence, the establishment of causation upon cohort analytic studies is, in the main, less liable to error than its establishment upon the results of case-control analytic studies. Accordingly, the continued development and refinement of methodologic standards for case-control studies becomes a high priority, especially in view of their increasingly frequent execution and appearance in the scientific literature.

RESEARCH PRIORITIES

On the basis of the foregoing, the following research priorities are nominated:

(1) The continued development of an annotated catalog of bias. Each citation should include a useful definition, a referenced example illustrating the magnitude and direction of its effects, and a description of the appropriate preventive measures, if any. I volunteer for this task, would welcome collaboration, and would appreciate receiving nominations and examples of additional biases.

(2) The empiric elucidation of the dynamics and results of these biases. Methodologists have too long ignored their responsibility to measure the occurrence and magnitude of bias, as shown in the 30 yr which elapsed between the description of the admission rate bias [9] and its first empiric demonstration [10, 3]. We are justly criticized for this lapse, and need to get to work.

(3) The development of methodologic standards for case-control studies. Such standards already exist for randomized trials of therapy and prevention [e.g. 25, 26]. The increasing frequency of case-control studies and their performance by an even wider group of investigators makes this a high priority [27]. The failure to respond here may lead to the publication of a rash of ill-conceived, seriously flawed case-control studies and a subsequent rejection of the entire approach by an inflamed scientific community.

(4) The validation of the proper role of case-control studies in clinical and health care decision-making. Sartwell has suggested that they are ill-suited for the evaluation of either the therapy or prophylaxis of disease, nor in his opinion should they be used to study diseases of high incidence and short duration [11]; others, including this author, have publicly questioned whether they should ever be used to make broad clinical policy without additional evidence from cohort analytic studies. Rather than seek to answer this question through rhetoric or anecdote, why not systematically study those questions in human biology about which both analytic and experimental evidence are available (as seen in the example of physical activity and recurrent myocardial infarction), identify agreements and disagreements, and quantitate the ability of the case-control study to predict the results of the proper randomized trial? The result might be 'bad theater', but it certainly would help to identify the proper place of the case-control study in the investigation of human health and disease.

REFERENCES

1. Slone D, Shapiro S, Miettinen O: Case control surveillance of serious illnesses attributable to ambulatory drug use. In: Epidemiological evaluation of drugs. **Proc Int Symp Epidemiological Evaluation of Drugs.** Milan, Italy, May 2–4 1977. Colombo F, Shapiro S, Slone D, Tognoni G (Eds) Amsterdam: Elsevier/ North Holland Biomedical Press, pp. 59–70

2. Fletcher R, Fletcher SW: Research architecture in general medical journals. Annual meeting of the Sydenham Society, 1978
3. Roberts RS, Spitzer WO, Delmore T, Sackett DL: An empirical demonstration of Berkson's bias. **J Chron Dis** 31: 119–128
4. Good IJ: A classification of fallacious arguments and interpretations. **Technometrics** 4: 125–132, 1962
5. Murphy EA: **The Logic of Medicine.** Baltimore: Johns Hopkins University Press, 1976
6. Feinstein AR: **Clinical Judgment.** Huntington: Krieger, 1967
7. Neyman J: Statistics—servant of all sciences. **Science** 122: 401, 1955
8. Friedman GD, Kannel WB, Dawber TR, McNamara PM: Comparison of prevalence, case history and incidence data in assessing the potency of risk factors in coronary heart disease. **Amer J Epid** 83: 366–378, 1966
9. Berkson J: Limitations of the application of fourfold table analysis to hospital data. **Biometrics Bull** 2: 47–53, 1946
10. Sackett DL: Discussion. In: The epidemiology venous thrombosis. **Milbank Mem Fund Quart** 50: 150, 1972
11. Sartwell PE: Retrospective studies: a review for the clinician. **Ann Intern Med** 81: 381–386, 1974
12. Horwitz RI, Feinstein AR: New methods of sampling and analysis to remove bias in case-control research. **Clin Res** 25: 459A, 1977
13. Miettinen OS: Matching and design efficiency in retrospective studies. **Amer J Epid** 91: 111–118, 1970
14. Seltzer CC, Bosse R, Garvey AJ: Mail response by smoking status. **Amer J Epid** 100: 453–477, 1974
15. Morris JN, Crawford MD: Coronary heart disease and physical activity of work. **Brit Med J** 4: 1485–1496, 1958
16. Rechnitzer PA, Pickard HA, Paivio AV, Yuhasz MS, Cunningham D: Long-term follow-up study of survival and recurrence rates following myocardial infarction in exercising and control subjects. **Circulation** 45: 853–857, 1972
17. Rechnitzer PA *et al.*: A controlled prospective study of the effect of endurance training on the recurrence rate of myocardial infarction. **Abstracts of the Annual Meeting of the Royal College of Physicians of Canada,** 1978
18. Fox AJ, White GC: Bladder cancer in rubber workers: do screening doctors' awareness distort the statistics? **Lancet** 1: 1009–1011, 1976
19. Nishiyama RH, Schmidt RW, Batsakis JG: Carcinoma of the thyroid gland in children and adolescents. **J Amer Med Assoc** 181: 1034–1038, 1962
20. Raventos A, Horn RC Jr, Ravdin IS: Carcinoma of the thyroid gland in youth: a second look ten years later. **J Clin Endocr Metab** 22: 886–891, 1962
21. Klemetti A, Saxen L: Prospective versus retrospective approach in the search for environmental causes of malformations. **Amer J Publ Hlth** 57: 2071–2075, 1967
22. Schull WJ, Cobb S: The intrafamilial transmission of rheumatoid arthritis. **J Chron Dis** 22: 217–222, 1969
23. Cochran WG: The planning of observational studies of human populations. **J Roy Stat Soc (Series A)** 128: 234–266, 1965
24. Brown GW: Berkson fallacy revisited. **Amer J Dis Child** 130: 56–60, 1976
25. Sackett DL: Design, measurement and evaluation in clinical trials. In: **Platelets, Drugs and Thrombosis.** Hirsch J *et al.* (Ed). Basel: S. Karger, 1975
26. Sackett DL: Periodic examination of patients at risk. In: **Cancer Epidemiology and Prevention.** Schottenfeld D (Ed). Springfield: Charles C. Thomas, 1975
27. Jick H, Vessey MP: Case-control studies in the evaluation of drug-induced illness. **Amer J Epid** 107: 1–7, 1978

APPENDIX. A CATALOG OF BIASES

(*A*) *Definition of bias*

'Any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth'. (Adapted from Murphy, **The Logic of Medicine,** Baltimore: John Hopkins University Press, 1976.)

(*B*) *Stages of research in which bias can occur*

(An outline of the catalog)
(1) In *reading-up* on the field.
(2) In specifying and selecting the study *sample.*
(3) In executing the experimental *manoeuvre* (or exposure).
(4) In *measuring* exposures and outcomes.
(5) In *analyzing* the data.
(6) In *interpreting* the analysis.
(7) In *publishing* the results [and back to (1)].

(*C*) *The catalog*

(Each bias is defined and followed by an example.)

(1) In *reading-up* on the field:
(a) *The biases of rhetoric.* Any of several tecchniques used to convince the reader without appealing to reason, e.g. Good IJ: a classification of fallacious arguments and interpretations. **Technometrics** 4: 125–132, 1962

(b) *The all's well literature bias.* Scientific or professional societies may publish reports or editorials which omit or play down controversies or disparate results, e.g. the debate on 'control' and the complications of diabetes, well shown in editorials in the **New Engl J Med** 294: 1004, 1976 and 296: 1228–1229, 1977

(c) *One-sided reference bias.* Authors may restrict their references to only those works that support their position; a literature review with a single starting point risks confinement to a single side of the issue, e.g. Platt and Pickering on the inheritance of hypertension; Hamilton, Pickering *et al.*: **Clin Sci** 24: 91–108, 1963; Platt: **Lancet** 1: 899–904, 1963

(d) *Positive results bias.* Authors are more likely to submit, and editors accept, positive than null results, e.g. multiple personal experiences

(e) *Hot stuff bias.* When a topic is hot, neither investigators nor editors may be able to resist the temptation to publish additional results, no matter how preliminary or shaky, e.g. recent publications concerning medication compliance

(2) In specifying and selecting the study sample:

(a) *Popularity bias.* The admission of patients to some practices, institutions or procedures (surgery, autopsy) is influenced by the interest stirred up by the presenting condition and its possible causes, e.g. White: **Brit Med J** 2: 1284–1288, 1953

(b) *Centripetal bias.* The reputations of certain clinicians and institutions cause individuals with specific disorders or exposures to gravitate toward them, e.g. the striking rate of posterior fossa cerebral aneurysms reported from the University of Western Ontario

(c) *Referral filter bias.* As a group of ill are referred from primary to secondary to tertiary care, the concentration of rare causes, multiple diagnoses and 'hopeless cases' may increase, e.g. secondary hypertension at the Cleveland Clinic; Gifford: **Milbank Mem Fund Quart** 47: 170–186, 1969

(d) *Diagnostic access bias.* Individuals differ in their geographic, temporal and economic access to the diagnostic procedures which label them as having a given disease, e.g. Andersen, Andersen: Patterns of use of health services. In: **Handbook of Medical Sociology.** Freeman *et al.* (Ed). Englewood Cliffs: Prentice-Hall, 1972

(e) *Diagnostic suspicion bias.* A knowledge of the subject's prior exposure to a putative cause (ethnicity, taking a certain drug, having a second disorder, being exposed in an epidemic) may influence both the intensity and the outcome of the diagnostic process, e.g. the possibility that rubber workers were victims of this bias was studied by Fox, White: **Lancet** 1: 1009–1010, 1976

(f) *Unmasking (detection signal) bias.* An innocent exposure may become suspect if, rather than causing a disease, it causes a sign or symptom which precipitates a search for the disease, e.g. the current controversy over post-menopausal estrogens and cancer of the endometrium

(g) *Mimicry bias.* An innocent exposure may become suspect if, rather than causing a disease, it causes a (benign) disorder which resembles the disease, e.g. Morrison *et al.*: **Lancet** 1: 1142–1143, 1977

(h) *Previous opinion bias.* The tactics and results of a previous diagnostic process on a patient, if known, may affect the tactics and results of a subsequent diagnostic process on the same patient, e.g. multiple personal experiences with referred hypertensive patients

(i) *Wrong sample size bias.* Samples which are too small can prove nothing; samples which are too large can prove anything

(j) *Admission rate (Berkson) bias.* If hospitalization rates differ for different exposure/disease groups, the relation between exposure and disease will become distorted in hospital-based studies. Berkson: **Biometrics Bull** 2: 47–53, 1946; Roberts RS, Spitzer WO, Delmore T, Sackett DL: **J Chron Dis** 31: 119–128

(k) *Prevalence–incidence (Neyman) bias.* A late look at those exposed (or affected) early will miss fatal and other short episodes, plus mild or 'silent' cases and cases in which evidence of exposure disappears with disease onset. Neyman: **Science** 122: 401, 1955

(l) *Diagnostic vogue bias.* The same illness may receive different diagnostic labels at different points in space or time, e.g. British 'bronchitis' versus North American 'emphysema'; Fletcher *et al.*: **Amer Rev Resp Dis** 90: 1–13, 1964

(m) *Diagnostic purity bias.* When 'pure' diagnostic groups exclude co-morbidity they may become non-representative

(n) *Procedure selection bias.* Certain clinical procedures may be preferentially offered to those who are poor risks, e.g. selection of patients for 'medical' versus 'surgical' therapy; Feinstein: **Clin Biostatistics** 76, 1977

(o) *Missing clinical data bias.* Missing clinical data may be missing because they are normal, negative, never measured, or measured but never recorded

(p) *Non-contemporaneous control bias.* Secular changes in definitions, exposures, diagnoses, diseases and treatments may render non-contemporaneous controls non-comparable, e.g. Feinstein: **Clin Biostatistics:** 89–104, 1977

(q) *Starting time bias.* The failure to identify a common starting time for exposure or illness may lead to systematic misclassification, e.g. Feinstein: **Clin Biostatistics:** 89–104, 1977

(r) *Unacceptable disease bias.* When disorders are socially unacceptable (V.D., suicide, insanity) they tend to be under-reported

(s) *Migrator bias.* Migrants may differ systematically from those who stay home, e.g. Krueger, Moriyama: **Amer J Publ Hlth** 57: 496–503, 1967

(t) *Membership bias.* Membership in a group (the employed, joggers, etc.) may imply a degree of health which differs systematically from that of the general population, e.g. exercise and recurrent myocardial infarction. Rechnitzer *et al.*: **Circulation** 45: 853–857, 1972 and **J Roy Coll Phys:** 29–30, 1978

(u) *Non-respondent bias.* Non-respondents (or 'late comers') from a specified sample may exhibit exposures or outcomes which differ from those of respondents (or 'early comers'), e.g. cigarette smokers; Seltzer *et al.*: **Amer J Epid** 100: 453–547, 1974

(v) *Volunteer bias.* Volunteers or 'early comers' from a specified sample may exhibit exposures or outcomes

(they tend to be healthier) which differ from those of non-volunteers or 'late comers', e.g. volunteers for screening; Shapiro *et al.*: **JAMA** 215: 1777–1785, 1971

(3) In executing the experimental manoeuvre (or exposure):

(a) *Contamination bias.* In an experiment when members of the control group inadvertently receive the experimental manoeuvre, the difference in outcomes between experimental and control patients may be systematically reduced, e.g. recent drug trials involving aspirin

(b) *Withdrawal bias.* Patients who are withdrawn from an experiment may differ systematically from those who remain, e.g. in a neurosurgical trial of surgical versus medical therapy of cerebrovascular disease, patients who died or stroked-out during surgery were withdrawn as 'unavailable for follow-up' and excluded from early analyses

(d) *Compliance bias.* In experiments requiring patient adherence to therapy, issues of efficacy become confounded with those of compliance, e.g. it is the high risk coronary patients who quit exercise programs; Oldridge *et al.*: **Canad Med Assoc J** 118: 361–364, 1978

(e) *Therapeutic personality bias.* When treatment is not 'blind', the therapist's convictions about efficacy may systematically influence both outcomes (positive personality) and their measurement (desire for positive results)

(f) *Bogus control bias.* When patients who are allocated to an experimental manoeuvre die or sicken before or during its administration and are omitted or re-allocated to the control group, the experimental manoeuvre will appear spuriously superior

(4) In measuring exposures and outcomes:

(a) *Insensitive measure bias.* When outcome measures are incapable of detecting clinically significant changes or differences, Type II errors occur

(b) *Underlying cause bias* (*rumination bias*). Cases may ruminate about possible causes for their illnesses and thus exhibit different recall or prior exposures than controls, e.g. Sartwell: **Ann Int Med** 81: 381–386, 1974 (see also the Recall bias)

(c) *End-digit preference bias.* In converting analog to digital data, observers may record some terminal digits with an unusual frequency, e.g. a notorious problem in the measurement of blood pressure; Rose *et al.*: **Lancet** 1: 296–300, 1964

(d) *Apprehension bias.* Certain measures (pulse, blood pressure) may alter systematically from their usual levels when the subject is apprehensive, e.g. blood pressure during medical interviews; McKegney, Williams: **Amer J Psychiat** 123: 1539–1545, 1967

(e) *Unacceptability bias.* Measurements which hurt, embarrass or invade privacy may be systematically refused or evaded

(f) *Obsequiousness bias.* Subjects may systematically alter questionnaire responses in the direction they perceive desired by the investigator

(g) *Expectation bias.* Observers may systematically err in measuring and recording observations so that they concur with prior expectations, e.g. house officers tend to report 'normal' fetal heart rates: Day *et al.*: **Brit Med J** 4: 422–424, 1968

(h) *Substitution game.* The substitution of a risk factor which has not been established as causal for its associated outcome. Yerushalmy: In: **Controversy in Internal Medicine.** Ingelfinger *et al.* (Eds). 1966

(i) *Family information bias.* The flow of family information about exposure and illness is stimulated by, and directed to, a new case in its midst, e.g. different family histories of arthritis from affected and unaffected sibs; Schull, Cobb: **J Chron Dis** 22: 217–222, 1969

(j) *Exposure suspicion bias.* A knowledge of the subject's disease status may influence both the intensity and outcome of a search for exposure to the putative cause, e.g. Sartwell: **Ann Int Med** 81: 381–386, 1974

(k) *Recall bias.* Questions about specific exposures may be asked several times of cases but only once of controls. (See also the underlying cause bias)

(l) *Attention bias.* Study subjects may systematically alter their behavior when they know they are being observed, e.g. Hawthorne revisited

(m) *Instrument bias.* Defects in the calibration or maintenance of measurement instruments may lead to systematic deviations from true values

(5) In analyzing the data:

(a) *Post-hoc significance bias.* When decision levels or 'tails' for α and β are selected *after* the data have been examined, conclusions may be biased

(b) *Data dredging bias* (*looking for the pony*). When data are reviewed for all possible associations without prior hypothesis, the results are suitable for hypothesis-forming activities only

(c) *Scale degradation bias.* The degradation and collapsing of measurement scales tends to obscure differences between groups under comparison

(d) *Tidying-up bias.* The exclusion of outlyers or other untidy results cannot be justified on statistical grounds and may lead to bias, e.g. Murphy: **The Logic of Medicine:** p. 250, 1976

(e) *Repeated peeks bias.* Repeated peeks at accumulating data in a randomized trial are not dependent, and may lead to inappropriate termination

(6) In interpreting the analysis:

(a) *Mistaken identity bias.* In compliance trials, strategies directed toward improving the patient's compliance may, instead or in addition, cause the treating clinician to prescribe more vigorously; the effect upon achievement of the treatment goal may be misinterpreted, e.g. Sackett: Priorities and methods for future research. In: **Compliance with Therapeutic Regimens.** Sackett DL, Haynes RB (Eds). 1976

(b) *Cognitive dissonance bias.* The belief in a given mechanism may increase rather than decrease in the face of contradictory evidence, e.g. Sackett: How can we improve patient compliance? In: **Controversies in Therapeutics.** Lasagna L (Ed). In press

(c) *Magnitude bias.* In interpreting a finding the selection of a scale of measurement may markedly affect the interpretation, e.g. $1,000,000 may also be 0.0003% of the national budget; Murphy: **The Logic of Medicine:** p. 249, 1976

(d) *Significance bias.* The confusion of statistical significance, on the one hand, with biologic or clinical or health care significance, on the other hand, can lead to fruitless studies and useless conclusions, e.g. Feinstein: **Clin Biostatistics:** p. 258, 1977

(e) *Correlation bias.* Equating correlation with causation leads to errors of both kinds, e.g. Hill: **Principles of Medical Statistics.** 9th ed. pp. 309–320, 1971

(f) *Under-exhaustion bias.* The failure to exhaust the hypothesis space may lead to authoritarian rather than authoritative interpretation, e.g. Murphy: **The Logic of Medicine:** p. 258, 1976