

# CLINICAL EPIDEMIOLOGY ROUNDS

## How to read clinical journals: IV. To determine etiology or causation

DEPARTMENT OF CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS,  
McMASTER UNIVERSITY HEALTH SCIENCES CENTRE

This is the fourth in the current consecutive series of Clinical Epidemiology Rounds devoted to efficient strategies and tactics for reading clinical journals. The first four guides for reading a clinical journal apply to any article (consider the title, the authors, the summary and the site) and appear in Fig. 1. The fifth guide depends on why the article is being read; the reason that will be considered in this round is to learn more about the etiology and causation of human illness.

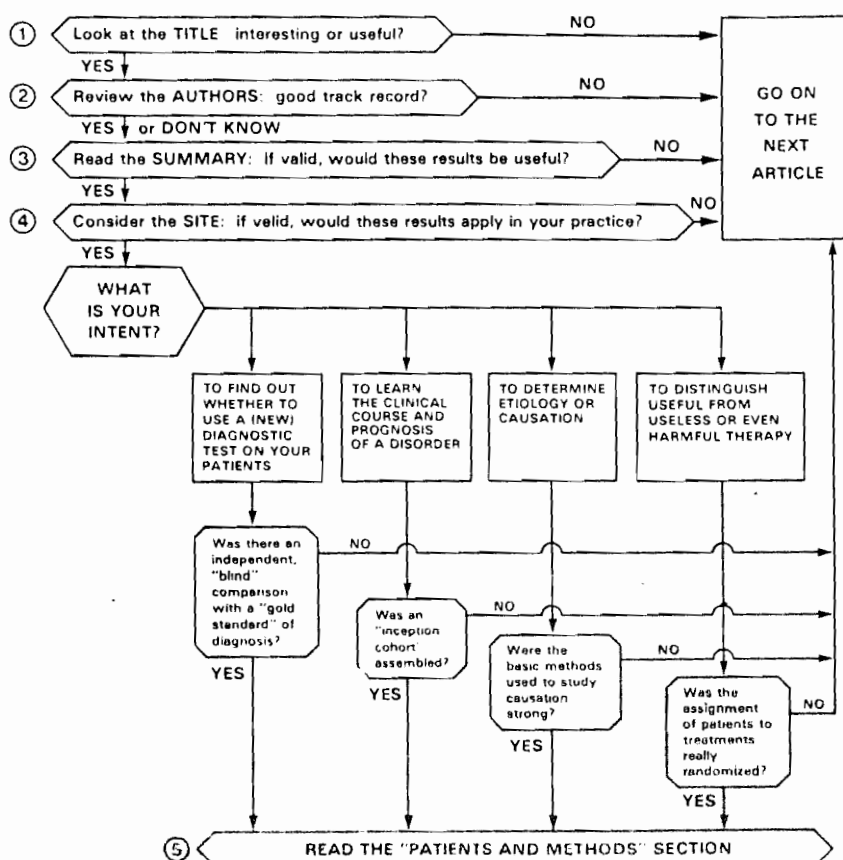
Perusal of a grab sample of issues from volume 121 of the Journal reveals that its clinical readers are faced with claims about the etiology and causation of human illness every time they read. For example, they are warned about dietary fibre and colon cancer;<sup>1</sup> cimetidine is implicated as a cause for diminished libido;<sup>2</sup> it is proposed that cigarette smoking has an effect on the risk of occupational lung cancer;<sup>3</sup> heavy tea consumption is blamed for iron deficiency;<sup>4</sup> they learn about drugs that cause dependence;<sup>5</sup> viral encephalitis is proposed as a cause of Huntington's chorea;<sup>6</sup> they are told that malfunctioning brown fat may be a cause of obesity;<sup>7</sup> *Campylobacter* is implicated in the etiology of ileocolitis;<sup>8</sup> they are asked to reconsider whether colour-blindness in auto drivers causes traffic acci-

dents;<sup>9</sup> and, finally, the roles of cold snaps and snow-shovelling as causes of sudden death are debated.<sup>10</sup>

These 10 "cases in causation" have important implications for both private and public practitioners because some powerful recommendations are made: we should try to

change our patients' eating habits, have second thoughts about prescribing some specific drugs, discourage tea-drinking in certain native groups, consider (or reconsider) stopping colour-blind people from driving, and stop lots of patients from shovelling snow.

When we add to the foregoing the



Reprint requests to: Prof. Kilgore S. Trout, Rm. 3V43F, McMaster University Health Sciences Centre, 1200 Main St., W. Hamilton Ont L8N 3Z5

FIG. 1—The first steps in how to read articles in a clinical journal.

bevy of claims for causation that our patients, to their distress, come upon in the lay press and on television, it becomes clear that clinicians are forced to make judgements and to give advice about causation all the time.

To help meet these demands for instant sagacity we have brought together some "applied principles of common sense" that should help the busy clinician assess an article that claims to show causation. They are distilled from the work of a number of methodologists, most notably Austin Bradford Hill.<sup>11</sup>

The application of these common-sense principles involves two steps. First, readers should scan the Methods section of the article to see whether the basic methods used were strong or weak. Second, they should then apply a set of "diagnostic tests" for causation to the remainder of the article.

#### Step one: Deciding whether the basic methods used were strong or weak

Sometimes you can identify the basic method used in a study from its title; other times you must examine its abstract or Methods section. Thus, step one can be accomplished quickly, without having to read the Introduction or Discussion. This step is summarized in Table I.

Suppose we really wanted to find out whether snow-shovelling was a cause for heart attack in middle-aged (your age plus 5 years) men. What would be the most powerful sort of study we could find in the clinical literature?

Most of you, we hope, would start by looking for a true experiment in humans — a study in which middle-aged men were randomly allocated (by a system analogous to tossing a coin) to habitually shovel or not shovel snow each winter,\* and were then followed to see how many in each group died suddenly. Evidence from such a *randomized trial* is the soundest evidence we can

\*Those who balk at the feasibility of this approach should recognize that the point at issue here is validity, not feasibility. On the other hand, the authors could have provided the controls with snow blowers!

ever obtain about causation (whether it concerns etiology, therapeutics or any other causal issue), and the reasons for this, if not already clear, will become apparent as we proceed. The basic architecture of the randomized trial is shown in Table II.

Although the true experiment (randomized trial) gives us the most accurate (or valid) answer to a question of causation, and therefore represents the strongest method, we will not find it very often in our clinical reading. In many cases (including the present example) it is not feasible to do a randomized trial to determine etiology, and in some it is downright unethical. For example, who would ever consider carrying out a true experiment that would

deliberately cause viral encephalitis in a random half of a group of individuals to see whether they were rendered more likely to develop Huntington's chorea?<sup>12</sup>

Thus, we are much more likely to encounter the following subexperimental studies of the risk of heart attack from snow-shovelling. For example, the next most powerful study method, the *cohort study*, would identify two groups (or cohorts) of middle-aged men, one cohort that did and the other that did not shovel snow each winter. The investigators would then follow these two cohorts, counting the heart attacks that occurred in each. In this case the direction of inquiry is forward in time, as depicted in Table III. If the heart attack rate was higher in the cohort that shovelled snow, this would constitute reasonably strong evidence that snow-shovelling precipitated heart attacks. However, the strength of such a cohort analytic study is not as great as that of a randomized trial; the reason for this difference in strength is apparent if we consider the middle-aged man with

Table I—Step one: Deciding whether the basic methods used were strong or weak

Strength	Method
Strongest	Randomized clinical trial
	Cohort study
	Case-control study
Weakest	Case series

Table II—Basic structure of a randomized trial

		Outcome (heart attack)	
		Yes	No
Ⓡ	Exposed (shovelled snow)	a	b
	Not exposed (did not shovel snow)	c	d

Direction of inquiry →

Table III—Basic structure of a cohort study

		Outcome (heart attack)	
		Yes	No
?	Exposed*	a	b
	Not exposed*	c	d

Direction of inquiry →

\*Definitions as in Table II.

angina pectoris. First, is he more likely than his angina-free neighbour to avoid snow-shovelling or other activities that precipitate angina? Yes. Second, is he at higher risk than his neighbour of heart attack? Yes again. Thus, the cohort analytic study could provide a distorted answer to the causal question if men at high risk of heart attack for extraneous reasons\* were not equally distributed between the cohorts of those who did and did not shovel snow. We see, then, that we must view a subexperimental study such as the cohort analytic study with some caution and suspicion.

A second type of subexperimental study deserves even greater caution in interpretation — the *case-control study*. In a case-control study the investigator gathers "cases" of men who have suffered a heart attack and a "control" series of men who have not had a heart attack. Both groups of men are then questioned about whether they regularly shovel snow each winter. If those who had heart attacks were more likely to regularly shovel snow, this would constitute some evidence, though not very strong, that snow-shovelling might cause, or at least precipitate, heart attack. Thus, in this case the direction of inquiry is backwards in time, as shown in Table IV.

\*You may come upon the term *confounder* in your reading, and that's what angina is in this example. First, it is extraneous to the question posed (What are the effects of snow-shovelling?); second, it is a determinant of outcome (heart attack); and, finally, it is unequally distributed between the cohorts of exposed and nonexposed persons.

Why is the case-control analytic study low on the scale of strength? Because it is so very liable to bias. The case-control study is susceptible not only to bias from the angina patient we noted in the cohort study, but also to several other sorts of bias.<sup>12</sup> For example, if snow-shovelling precipitated not only heart attack but also sudden death, many victims would not survive long enough even to be included in a case-control study, much less to be interviewed. As a result, snow-shovelling would appear to be a benign pastime when, in fact, it was lethal for some middle-aged men. For this and other similar reasons, the results of case-control studies are tenuous at best in sorting out the etiology and causation of human illness.

One final type of subexperimental study deserves mention. This is the *case series*, in which an investigator might simply report that 60% of the men who had a heart attack were shovelling snow just before the onset of their infarcts. No comparison group is provided, and about all the reader can conclude is that heart attack *can* (but not necessarily *does*) follow snow-shovelling. Such case series, though often thought-provoking, are prone to overinterpretation, especially by their investigators. In terms of strength, case series are best used to stimulate other, more powerful investigations. All too often, however, they provoke authoritarian (rather than authoritative) clinical advice about etiology, prevention and therapy.

In summary, then, readers of

reports purporting to show etiology or causation should begin by deciding whether the basic methods used were strong or weak (Table I). If the basic method was a randomized trial, it is the strongest and usually can be trusted. This is how, for example, the best evidence was obtained on the real side effects produced by frequently used antihypertensive drugs.<sup>13</sup> A cohort analytic study, although weaker than a randomized trial, is always preferred to a case-control study, and can sometimes be trusted. Thus, the most convincing (but none the less disputed) evidence about the possible side effects of oral contraceptives comes from a large cohort study carried out by British general practitioners.<sup>14</sup> The case-control study is a weak design and has often led to erroneous conclusions (such as the now discredited link between reserpine and breast cancer<sup>15</sup>); however, for some extremely rare disorders (especially rare adverse drug reactions) we may have only case-control studies to go by and may be forced, however reluctantly, to trust them. Finally, it is not possible to tell whether any given case series can, all by itself, be trusted on an issue of etiology or causation. Thus, if other, stronger evidence is available, such case series should be passed over.

## Step two: Applying the diagnostic tests for causation

Having decided from the foregoing that the article warrants further consideration, readers should then turn to the Results, the Introduction and the Discussion to see how the data fit some common-sense rules of evidence. In making this causal decision, information should be sought relative to the diagnostic tests listed in Table V. They are discussed in order of decreasing importance, and we have suggested their impact upon the causal decision in Table VI.

The rules for interpreting clinical diagnostic tests that we described in an earlier round in this series (part II) can be applied here as well. For example, some of the tests for causation (such as evidence from randomized trials) are more accu-

Table IV—Basic structure of a case-control study

	Outcome (heart attack)	
	Cases	Controls
Exposed*	a	b
Not exposed*	c	d

← Direction of inquiry

\*Definitions as in Table II.

rate than others (such as analogy). Furthermore, many of them (such as temporality) are better for "ruling out" than for "ruling in" causation. Finally, epidemiologic sense and biologic sense, although prominent in many articles, are low on the list because they have relatively low specificity; it is possible to "explain" almost any set of observations.

### 1. Is there evidence from true experiments in humans?

As we explained earlier, these are investigations in which identical groups of individuals, generated through random allocation, are or are not exposed to the putative causal factor and are followed for the occurrence of the outcome of interest.

As we have just seen, this is the best evidence we will ever have, but it is not always available and is rarely the initial evidence for causation. None the less, any consideration of an issue of causation should begin with a search for a randomized trial.

### 2. Is the association strong?

Strength here means that the odds favour the outcome of interest with, as opposed to without, exposure to the putative cause; the higher the odds, the greater the strength.

There are different strategies for estimating the strength of an association. In the randomized trial and cohort study (Tables II and III) patients who are or are not exposed to the putative cause are carefully followed up to find out whether the adverse reaction or outcome

develops. Such a cohort study would, for example, compare the occurrence of impotence among ulcer patients who received cimetidine and those who did not.<sup>2</sup>

Cohort studies (Table III) are methodologically attractive because, like randomized trials, they permit direct calculations of strength (relative risk) by comparing outcome rates in exposed and nonexposed persons as follows:

$$\left( \frac{a}{a+b} \right) / \left( \frac{c}{c+d} \right)$$

However, as we learned in the previous section, cohort studies are often lengthy and expensive. Accordingly, the greater speed and lower cost of the case-control study (Table IV), in which patients with or without the outcome of interest (e.g., impotence) are selected and tracked backwards to their exposure to the putative cause (e.g., cimetidine), make it a much more popular approach, particularly as the first step in probing the conclusions of initial case series. Case-control or "trohoc"<sup>10</sup> studies pay a methodologic price for their savings in time and dollars. Strength or relative risk can only be indirectly estimated, from  $ad/bc$ . This calculation, though justified algebraically, is viewed with some scepticism.<sup>16</sup>

Moreover, as we have seen, case-control studies are particularly vulnerable to a series of systematic

distortions (biases) that may lead to erroneous estimates of the strength of association and, therefore, incorrect conclusions about causation. Some of these biases were discussed in a previous round in this series (part III), and still others are described in detail elsewhere for readers who want to pursue this.<sup>12</sup>

A review of the potential effects of these biases in distorting the conclusions of case-control and cohort studies leads to two conclusions. First, case-control studies are subject to more sources of bias than are cohort studies. Second, whereas one can usually anticipate and overcome (through appropriate and rigorously applied methods) the biases affecting cohort studies, this solution is either much more difficult or impossible in the case-control strategy. As a result, readers can place considerable confidence in estimates of strength from a randomized trial, fair confidence in an estimate of strength from a cohort study and only a little confidence in an estimate of strength from a case-control study.

### 3. Is the association consistent from study to study?

The repetitive demonstration by different investigators of an association between exposure to the putative cause and the outcome of interest, using different strategies and in different settings, constitutes consistency. Thus, much of the credibil-

Table V—Step two: Applying the diagnostic tests for causation\*

1. Is there evidence from true experiments in humans?
2. Is the association strong?
3. Is the association consistent from study to study?
4. Is the temporal relationship correct?
5. Is there a dose-response gradient?
6. Does the association make epidemiologic sense?
7. Does the association make biologic sense?
8. Is the association specific?
9. Is the association analogous to a previously proven causal association?

\*Listed in decreasing order of importance.

Table VI—Importance of individual diagnostic tests in making the causal decision

Diagnostic test*	Effect of test result on causal decision†		
	Test result consistent with causation	Test result neutral or inconclusive	Test result opposes causation
Human experiments	++++	---	----
Strength of association			
From randomized trial	++++	---	----
From cohort study	+++	--	---
From case-control study	+	0	-
Consistency	+++	--	----
Temporality	++	--	----
Gradient	++	-	--
Epidemiologic sense	++	-	--
Biologic sense	+	0	-
Specificity	+	0	-
Analogy	+	0	0

\*Listed in decreasing order of importance.

†+ = causation supported; - = causation rejected; 0 = causal decision not affected. The number of plus and minus signs indicates the relative contribution of the diagnostic test to the causal decision.

ty of the causal link between smoking and lung cancer arises from the repeated demonstration of a strong statistical association in case-control, cohort and other study designs.

#### 4. Is the temporal relationship correct?

A consistent sequence of events of exposure to the putative cause, followed by the occurrence of the outcome of interest, is required for a positive test of temporality. Although this diagnostic test looks easy to apply, it is not. What if a second predisposing factor or a very early stage of the disorder itself is responsible for both exposure to the putative causal factor and progression to the full-blown outcome? Indeed, such an explanation might apply to studies that have linked the use of illicit stimulant or depressant drugs to the subsequent diagnosis of psychosis or depression, respectively.<sup>17</sup> Did the different illicit drugs cause specific forms of subsequently diagnosed mental illness, or did individuals with different sub-clinical but progressive mental illness seek out the specific drugs? Understandably, this yardstick is easier to apply to cohort than to case-control studies, since the latter can imply a temporal association between "exposure" and "outcome" only after both have occurred.

#### 5. Is there a dose-response gradient?

The demonstration of increasing risk or severity of the outcome of interest in association with an increased "dose" or duration of exposure to the putative cause satisfies this diagnostic test. For example, in a report linking conjugated estrogens with endometrial carcinoma,<sup>18</sup> the relative risk of endometrial cancer rose from 5.6% among those who used the drug for 1 to 4.9 years to 7.2% among those who used it for 5 to 6.9 years and, finally, to 13.9% for those who used it for 7 or more years.

Reverse gradients are useful too. Indeed, some of the most compelling evidence of the link between cigarette smoking and lung cancer is the progressive decline in cancer

risk that has been reported as previous smokers celebrate anniversaries of their last cigarette.

#### 6. Does the association make epidemiologic sense?

This guide is met when the article's results are in agreement with our current understanding of the distributions of causes and outcomes in humans.

For example, Freeman,<sup>1</sup> reviewing the possible role of dietary fibre in the pathogenesis of colon cancer, noted several studies in which the distribution of dietary fibre among different geographic areas or populations was inversely related to the occurrence of colon cancer in the same areas and populations. Recognizing the tenuous nature of such epidemiologic correlations (after all, the declining birth rate in Europe has closely paralleled the disappearance of storks from its cities), Freeman called for "long-term prospective studies" to better define the role of dietary fibre in cancer in humans.

#### 7. Does the association make biologic sense?

Is there agreement with current understanding of the responses of cells, tissues, organs and organisms to stimuli? It is with this yardstick that nonhuman experimental data should be measured. Although virtually any set of observations can be made biologically plausible (given the ingenuity of the human mind and the vastness of the supply of contradictory biologic facts), some biologic observations can be compelling, such as Himms-Hagen's description<sup>7</sup> of the production of massive obesity in certain strains of mice whose brown fat had only a limited capacity for thermogenesis.

#### 8. Is the association specific?

The limitation of the association to a single putative cause and a single effect satisfies this diagnostic test. Examples here include some of the highly characteristic genetic disorders in which derangements in a single enzyme or another protein produce quite specific illnesses, such as hemophilia A or cystinuria. This is one of the minor diagnostic tests, being only moderately useful —

and, even then, only when the illness is present. The weakness of this test is underscored when you consider that teratogens commonly have multiple effects in several organ systems.

#### 9. Is the association analogous to a previously proven causal association?

The last and least of the diagnostic tests, this yardstick would link the scrotal cancer of chimney sweeps in a former era with the more recent appearance of lung cancer among persons who inhale, rather than wear, the products of combustion.

### Use of these guides to reading

When confronted by a question of causation, you can use these nine diagnostic tests to distil your clinical reading and, with the assistance of judgements such as those shown in Table VI, reach a causal conclusion. Even before reading, you can use these guides to increase the efficiency of a literature search, focusing attention on the publications that will shed the strongest light on the causal question and warn against accepting plausible but biased conclusions.

Even after extensive reading and the application of all nine diagnostic tests, however, you may remain uncertain about whether, for example, drug A really causes illness B. What do you do then, and how do you translate all of this deliberation into clinical action?

We suggest that this "decision for action" has two components (Fig. 2). First is our *certainty about causation*, which is based upon the results of applying the nine diagnostic tests for causation to our clinical reading. Second is our *consideration of the consequences of the alternative courses of action* open to us (recognizing that these courses of action include noninterference as well as maintenance of the status quo). The decision for

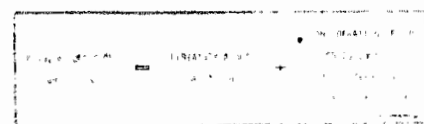


FIG. 2—Components of a "clinical decision for action".



action results from the interplay of these two components. Consider two examples:

The three reports that appeared abruptly in 1974 indicating reserpine as a cause of breast cancer<sup>19-21</sup> precipitated a crisis in the management of hypertension. How were we to advise and treat patients whose high blood pressure was kept under control with this drug? The first component of this decision considered the degree of certainty that reserpine did, indeed, cause breast cancer; it was never very great (in fact, the drug was later virtually pardoned by some of its earlier accusers<sup>22</sup>). On the other hand, the second component of this decision identified an alternative course of action that was highly attractive to many Canadian clinicians: switching appropriate patients from reserpine to propranolol. Thus, in this case even a low degree of certainty about causation was attended by the clinical decision to stop prescribing a drug for many patients because alternative treatment was available.

In contrast, the degree of certainty that oral contraceptives cause thromboembolism is much higher. None the less, oral contraceptives are still widely used. Although the reasoning behind the decision to continue oral contraceptive use in the face of growing evidence that it causes thromboembolism is complex, it is due, in part, to the second component of the decision: the consequences of alternative approaches to birth control may be judged even less desirable than the small but real risk of thromboembolism. Thus, the use of oral contraceptives continues (and, interestingly, the diagnostic test of the dose-response gradient is involved to justify the progressive reduction of certain hormonal constituents of oral contraceptives).

The diagnosis of causation is not simply arithmetical, and the strategies and tactics for making this judgement are still primitive. The diagnostic tests presented here are a start, and we suggest that their use, particularly when clearly specified before a review of relevant data, will lead to more rational — albeit less colourful — discussions of causation in medicine.

The next and final round in this series will address how to read clinical journals to distinguish useful from useless or even harmful therapy.

## References

1. FREEMAN HJ: Dietary fibre and colonic neoplasia. *Can Med Assoc J* 1979; 121: 291-296
2. BIRON P: Diminished libido with cimetidine therapy (C). *Ibid*: 404-405
3. CHOUVY AC: Occupational lung cancer and smoking: a review in the light of current theories of carcinogenesis. *Ibid*: 548-555
4. FARKAS CS: Body iron status associated with tea consumption (C). *Ibid*: 706
5. JACOB MS, STELLERS EM: Use of drugs with dependence liability. *Ibid*: 717-724
6. AVERBACK P: Viral encephalitic pathogenesis of Huntington's chorea? (C). *Ibid*: 1060-1062
7. HIMMS-HAGEN J: Obesity may be due to a malfunctioning of brown fat. *Ibid*: 1361-1364
8. LAMBERT JR, FISCHER ME, KARMAI MA, NEWMAN A: Campylobacter ileocolitis: an inflammatory bowel disease. *Ibid*: 1377-1379
9. ILES JDH: Colour-blind drivers of motor vehicles (C). *Ibid*: 1566
10. ANDERSON TW, ROCHARD C: Cold snaps, snowfall and sudden death from ischemic heart disease. *Ibid*: 1580-1583
11. HILL AB: *Principles of Medical Statistics*, 9th ed, Lancet, London, Engl. 1971: 312-320
12. SACKETT DL: Bias in analytic research. *J Chronic Dis* 1979; 32: 51-63
13. Veterans Administration Cooperative Study Group on Antihypertensive Agents: Effects of treatment on morbidity in hypertension: III. Influence of age, diastolic pressure, and prior cardiovascular disease; further analysis of side effects. *Circulation* 1972; 45: 991-1004
14. *Oral Contraceptives and Health. An Interim Report from the Oral Contraceptive Study of the Royal College of General Practitioners*, Pitman Med, London, Engl, 1974
15. LABARTHE DR: Methodologic variation in case-control studies on reserpine and breast cancer. *J Chronic Dis* 1979; 32: 95-104
16. FEINSTEIN AR: Clinical biostatistics. XX. The epidemiologic triad, the ablative risk ratio, and "retrospective" research. *Clin Pharmacol Ther* 1973; 14: 291-307
17. McLELLAN AT, WOODY GE, O'BRIEN CP: Development of psychiatric illness in drug abusers. Possible role of drug preference. *N Engl J Med* 1979; 301: 1310-1314

18. ZIEI HK, FINKEL WD: Increased risk of endometrial carcinoma among users of conjugated estrogens. *N Engl J Med* 1975; 293: 1167-1170
19. Boston Collaborative Drug Surveillance Program: Reserpine and breast cancer. *Lancet* 1974; 2: 669-671
20. ARMSTRONG B, STEVENS N, DOLL R: Retrospective study of the association between use of rauwolfia derivatives and breast cancer in English women. *Ibid*: 672-675
21. HEINONEN OP, SHAPIRO S, TUOMINEN L, TURUNEN MI: Reserpine use in relation to breast cancer. *Ibid*: 675-677
22. SHAPIRO S, SLOAN D: The case-control study: consensus and controversy. *Comment. J Chronic Dis* 1979; 32: 105-107

## BOOKS

This list is an acknowledgement of books received. It does not preclude review at a later date.

**ADVANCES IN HEART DISEASE.** Volume 3. Edited by Dean T. Mason. Cune & Stratton, New York; Academic Press Canada, Don Mills, Ont., 1980. \$87.10. ISBN 0-8089-1284-4

**AFTER THE EMERGENCY.** Follow-up Instructions for Patients in English and Spanish. David B. McMicken. Spanish translations by Frank Quintero. 167 pp. EM Books, New York, 1980. \$6.95, paperback

**BLOOD PRESSURE CONTROL.** Volume 1. Thomas G. Coleman. 248 pp. Eden Press Inc., Westmount, PQ, 1981. \$32.50. ISBN 0-88831-088-9

**CHILDREN IN NEED OF SPECIAL CARE.** Human Horizons Series. Thomas J. Weihs. 184 pp. Souvenir Press Ltd., London; John Wiley & Sons Canada, Limited, Rexdale, Ont., 1971. Reprinted 1980. Price not stated, paperback. ISBN 0-285-62003-7

**THE CLINICAL TRAINING OF DOCTORS.** An Essay of 1793. Philippe Pinel. Edited and translated, with an introductory essay by Dora B. Weiner. 102 pp. Illust. The Johns Hopkins University Press, Baltimore, Maryland, 1981. \$7.50, paperback. ISBN 0-8018-2448-6

**CPR AND EMERGENCY CARDIAC CARE: LOOKING TO THE FUTURE.** The Proceedings of the New York Health Association Conference on Current Practice and Current Research in Cardiopulmonary Resuscitation and Emergency Cardiac Care. Edited by Joseph Schulger and Alan F. Lyon. 144 pp. Illust. EM Books, New York, 1980. \$8.95, paperback

continued on page 103