Epi 607c
May/June '88

SPECIAL ARTICLE

CLINICAL RESEARCH IN GENERAL MEDICAL JOURNALS

A 30-Year Perspective

ROBERT H. FLETCHER, M.D., AND SUZANNE W. FLETCHER, M.D.

Abstract Little is known about the frequency with which various research designs appear in the clinical literature and how this frequency has changed in recent years. This study describes the research designs used in 612 articles randomly selected from original research published in three general medical journals from 1946 to 1976.

Cross-sectional studies increased from 25 to 44 per cent, cohort studies declined from 59 to 34 per cent, and clinical trials increased from 13 to 21 per cent of articles ($P<0.001$). Randomized controlled trials comprised 5 per cent of articles published in 1976 and were not represented 30 years before. In 1976, 37 per cent of articles reported on 10 subjects or less, and this number has not changed substantially since 1946.

The frequency of studies with weak research designs has increased in these general medical journals over the past 30 years. The trend deserves critical attention. (N Engl J Med 301:180-183, 1979)

CLINICAL research must provide sound data that clinicians can rely on in making decisions about patients. This kind of research should help to establish a body of knowledge that is useful in answering the following clinical questions: What medical condition does the patient have? How common is the condition? What are its causes? What is its prognosis? And how effective and risky is the treatment? Presumably, clinicians read original research published in medical journals to find answers to such questions.

For some diseases, answers to these questions were provided by great clinicians of the past. However, many of the questions that remain concern relations so subtle that answers are not obvious, even to the most experienced and astute physicians. This is particularly true of chronic conditions. To explore these relations rigorously, powerful clinical research designs have been developed. Among these designs are randomized controlled trials, which, when well conducted, are the least fallible means of assessing the efficacy of medical treatment[1,2]; cohort ("prospective") studies, in which the natural history of disease is observed over time and differences between subgroups are corrected for by matching or adjustment in situations where experimentation is not possible[3]; and case-control ("retrospective") studies, which provide an efficient way of exploring cause and effect relations for uncommon diseases.[4]

However, other research methods, with less sound scientific credentials, are frequently encountered in the clinical literature. Among these methods are case reports, uncontrolled clinical trials and various kinds of cross-sectional studies in which observations are made on subjects at essentially one point in the course of their illness. Such research is more prone to problems of chance, bias and confounding factors that can threaten the validity of conclusions based on the research.

Little has been written about the frequency with which readers of original research in general medical journals encounter the various kinds of research designs[5,6] or whether there has been a change in this frequency over time. Feinstein devised a detailed classification of research designs and used it to describe the articles appearing in two general medical journals during a recent six-month period.[5] He found that 61 per cent of the published research was cross sectional and that 39 per cent was longitudinal. The present study describes the frequency with which various research designs have appeared in the general medical literature over the past 30 years. In addition, we have looked at the kinds of variables used to describe the clinical phenomena studied.

METHOD

Three journals — *The Journal of the American Medical Association, The Lancet,* and *The New England Journal of Medicine* — were selected for review. These three were chosen because they are the most widely circulated English-language journals that present original research to a general clinical audience,[7] and because they are well respected for the manner in which they fulfill this role.

These journals were reviewed for the years 1946, 1956, 1966 and 1976. For each issue of the journals in these years, we selected one original article by use of a table of random numbers. The sampling fraction was about 20 per cent. Twelve issues, most of them published in 1946, did not contain original research and so could not be included in the sample. A few of the selected articles (3 per cent) were not considered to be clinical research, in that they did not even indirectly involve human subjects; these articles were replaced by another article in the same issue.

From each of 612 articles, the following information was obtained: the number of authors; the number of subjects; the longest period of time any of the subjects was studied; whether the article was commented on in an editorial in the same issue; whether

From the departments of Medicine and of Epidemiology and Health, McGill University, and the McGill University Clinic, Royal Victoria Hospital, Montreal, Quebec, Canada (address reprint requests to Dr. Robert H. Fletcher at the Robert Wood Johnson Clinical Scholars Program, Medical School, Wing B, Box 5, 207H, University of North Carolina, Chapel Hill, NC 27514).

the data apparently had been collected before or after the decision to perform the study; the variables used to describe the clinical phenomena studied (including symptoms, laboratory-test results, death, social or occupational function and emotional status); and whether there was explicit mention of cost (expressed in dollars or related units such as length of hospital stay) when results were presented.

In addition, information was abstracted that was used to classify the articles according to the following definitions of principal research designs:

Cross-sectional articles are those in which all observations on a given subject are made at essentially one point in time in the course of that subject's illness. Examples of this category include descriptions of the clinical presentations of disease, associations between clinical manifestations, comparisons of the relative value of diagnostic tests and various studies in which it is assumed during analysis of the data that one phenomenon precedes another even though the data are not actually collected over a period of time.

The remaining studies are "longitudinal," in that subjects are observed over a period of time. Longitudinal studies can be classified further according to the temporal direction in which subjects are pursued.

In *case–control* studies, subjects are pursued backward in time, from effects to possible causes. The term "case–control," also called "retrospective," was chosen because it is familiar to most clinicians. It is not an ideal term because not all studies proceeding backward in time, from subsequent to antecedent events, describe control groups and cases. Feinstein has suggested "trohoc" (cohort spelled backward) as an unambiguous term to describe studies in which subjects are pursued backward in time.[8]

In *cohort* research, subjects are followed forward in time, from purported causes to effects. Such research has also been referred to as "prospective." When patients experience interventions that were not contrived for the purposes of study but, rather, take place in the normal course of their illness, with or without treatment, the study is called a cohort study.

Clinical trials are cohort studies in which treatment is initiated specifically for evaluation and not just during the care or observation of patients. Such trials are further classified into *uncontrolled trials*, in which there is no concurrent comparison group; *nonrandom controlled trials*, in which concurrent comparison groups are allocated by means of some nonrandom process (e.g., convenience or clinical judgment); and *randomized controlled trials*, in which subjects are randomly allocated into treatment and control groups.

These working definitions of the principal research designs are mutually exclusive and describe most of the articles reviewed. We believe that these operational definitions conform reasonably well to accepted definitions of research structures as commonly used by clinicians. The reliability of the classification scheme was tested by means of a subsample of articles that were classified independently by both authors. There was strong agreement[9] between the two abstractors in the category of research structures chosen (unweighted kappa = 0.88). Most discordance was related to the distinction between cohort studies and uncontrolled trials, since it was not always clear whether treatment was begun primarily for patient care or research.

Articles reporting on 10 or fewer subjects were considered "case reports" and were analyzed separately. The cutoff point of 10 was chosen because it is usually not possible to exclude the role of chance in studies of this size. There was strong agreement (unweighted kappa = 0.64) between this definition of a case report and the presence or absence of the typical case-report format in which descriptions of different patients' clinical courses were included, usually preceded by "case 1, case 2..." and printed in small type.

The statistical significance of differences between years was calculated by use of analysis of variance for means, the chi-square test for proportions and the median test for medians.[10] All tests were two-tailed.

Results

Table 1 shows several characteristics of the articles published from 1946 to 1976. The number of authors per article has increased, particularly in the past decade ($P<0.001$). There has been no consistent change over the past 30 years in the number of subjects studied. There has been a trend toward shorter duration of follow-up for subjects in longitudinal studies, but this difference was not statistically significant. An increase was found in the proportion of articles based on data that had already been collected before the research was planned, as compared with data collected after the research was planned.

Figure 1 shows the relative frequencies of the various research designs in 1976. Forty-four per cent of all articles used a cross-sectional design. Half of this cross-sectional research (22 per cent of the total) was a description of the manifestations of disease, or associations between manifestations, at one time. One

Table 1. Selected Characteristics of Published Articles, 1946 to 1976.

Characteristic	1946 (151 articles)	1956 (149 articles)	1966 (157 articles)	1976 (155 articles)	P
No. of authors (mean ± SD)	2.0±1.4	2.3±1.6	2.8±1.2	4.9±7.3	<0.001
No. of subjects (median)	25	36	16	30	NS
Duration of follow-up for cohort studies (median no. of days)	232	300	152	182	NS
Data collection:					
After research was planned (%)	76	71	49	44	<0.05
Before research was planned (%)	24	29	51	56	

fifth (8 per cent of the articles from 1976) were evaluations of the properties of diagnostic tests, usually comparing them to some standard of accuracy and expressing the results in such terms as sensitivity, specificity and predictive value. Another fifth (8 per cent) would ordinarily be considered "case-control" research. Diseased and nondiseased subjects had been identified, and at the same time information had been elicited concerning their previous habits and exposures; it had then been assumed that these conditions actually did exist before the illness, although they had not been recorded in the past. None of the articles sampled in this year were truly "case-control," by the definition used in this study. Fifty-five per cent of all articles were longitudinal. Two thirds of these articles were cohort studies, and one third were clinical trials. Half the clinical trials had no concurrent control group. Five per cent of all studies reviewed were randomized controlled trials. A small proportion of the total (1 per cent) could not be classified.

Figure 2 shows the relative frequencies of the various research designs from 1946 to 1976. There has been a sharp fall in the proportion of longitudinal studies and a rise in cross-sectional studies, both occurring after 1956. Clinical trials have increased slightly during the study period. The change in the frequencies of these three kinds of studies from 1946 to 1976 was statistically significant ($P<0.01$). Randomized controlled trials have increased but have never comprised more than 5 per cent of the total.

The proportion of articles that received comment in an editorial in the same issue increased from 5 per cent in 1946 to 14 per cent in 1976 ($P<0.001$). The relative frequencies of the various research designs used in the articles singled out for editorial comment were similar to those for all articles in the same year. There was no striking difference among the three journals in the proportions of the various designs, and none of the differences observed were statistically significant. However, the number of articles sampled was too small to be reasonably certain that no such difference has existed.

Of the 612 articles, 234 (38 per cent) presented data on 10 or fewer subjects and can be considered "case reports." Seventy-eight of them (13 per cent) were reports of single cases. There was no consistent change in the frequency of case reports involving either single or several cases. Most case reports were longitudinal, describing the clinical course of selected patients. However, the proportion of longitudinal case reports declined from 90 to 64 per cent during the period of study, whereas cross-sectional case reports increased from 10 to 36 per cent ($P<0.001$). Virtually all the cross-sectional case reports were descriptions of the manifestations of disease for selected patients at one point in the course of their illness.

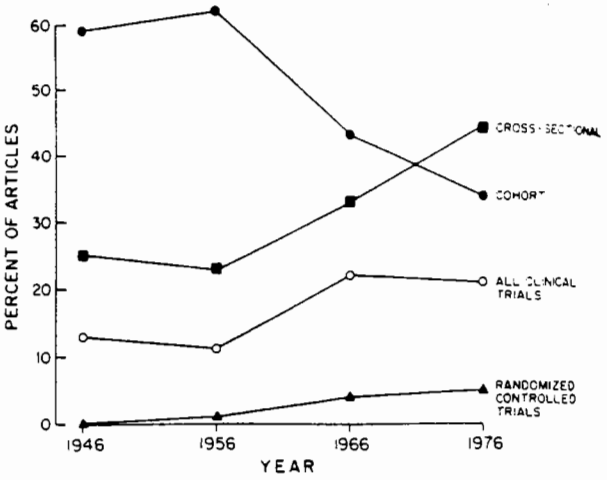

Figure 2. Research Designs for 612 Articles Published from 1946 to 1976.

Table 2 shows the variables used to describe the clinical phenomena under investigation. Laboratory tests were the most frequently reported clinical tool in every year, except 1946, and use of such measures has tended to increase with time. Clinical signs and symptoms have also been reported often, although with significantly less frequency in recent years ($P<0.05$). Similarly, the frequency with which the social, occupational or emotional status of subjects has been reported has diminished with time ($P<0.05$). Cost was explicitly mentioned in less than 2 per cent of the articles in any year.

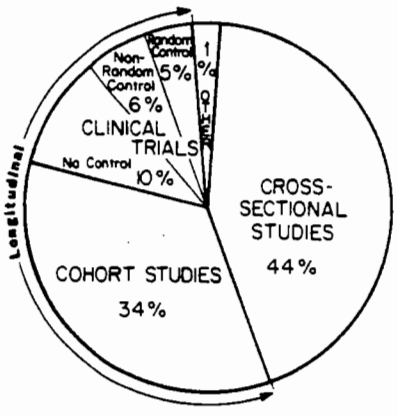

Figure 1. Research Designs for 155 Articles Published in 1976.

Discussion

Taken together, the data in this study suggest that the predominant research designs used in 1976 were less accurate ways of conducting clinical research than had been employed in earlier years. Longitudinal studies have declined in frequency, whereas cross-sectional studies have increased. It is more common for researchers to have used existing data rather than to have collected the data themselves, specifically to answer their research question. Neither the number of subjects studied nor the duration of follow-up observation for longitudinal studies has increased despite the advent of computer technology, which has made it easier to manage large amounts of data. Finally, the observations made on subjects seem to have become less relevant to clinical events; symptoms, death,

Table 2. Variables Used to Describe Clinical Phenomena in Articles Published from 1946 to 1976.

VARIABLE	PER CENT OF ARTICLES PER YEAR			
	1946 (151 ARTICLES)	1956 (149 ARTICLES)	1966 (157 ARTICLES)	1976 (155 ARTICLES)
Symptoms*	88	79	61	63
Clinical signs*	90	79	64	68
Laboratory tests	79	80	90	90
Death	26	26	20	18
Social or occupational function*	10	6	2	3
Emotional status*	7	8	5	1

*$P<0.05$ for differences among years.

social and occupational function and emotional status were less frequently assessed than they had been previously, whereas laboratory tests are now the most frequently reported kind of result. Similar trends away from patient-oriented research were found by Feinstein in his study of abstracts for the "Atlantic City meetings," of 1953 to 1969.[11,12]

There are probably many reasons why cross-sectional research and case reports comprise such a large part of published research, have done so for the past 30 years and have not declined in frequency in recent years. Case reports serve some indispensable roles. They are virtually the only means of surveillance for diseases of low frequency and the only way of studying such diseases. Similarly, cross-sectional studies are an appropriate way of studying clinically useful phenomena, such as the accuracy of diagnostic tests. However, both cross-sectional studies and case reports are unreliable when used as surrogates for longitudinal studies of larger numbers of subjects in exploring the etiology and natural history of diseases or the effects of treatment.

There are many examples in the medical literature where conclusions based on relatively weak research designs have been misleading and have been corrected by subsequent studies of stronger design. For example, uncontrolled trials suggested that mammary-artery ligation was extremely effective in relieving angina pectoris[13]; subsequently, controlled trials of the procedure showed that the effectiveness of surgical intervention could not be attributed to ligating the mammary artery.[14] Moreover, reviews of the literature on specific treatments (e.g., corticosteroids for bacterial infections,[15] anticoagulation after acute myocardial infarction[16] and coronary-artery-bypass grafts for coronary-artery disease[17]) have shown an association between the conclusions of the studies and the strength of the research designs used to establish these conclusions. This association supports the belief that studies based on weak designs can be misleading and should receive less weight when the literature on a subject is summarized.

Perhaps some of the trends we observed result from factors that are not directly related to the relative scientific merits of alternative designs. For example, the approach to clinical research might be influenced by rapid developments in technology, the mobility of both researchers and patients and academic competition ("publish or perish"). Longitudinal studies, for all their scientific advantages, are often expensive, slow and difficult to conduct. It may be that researchers believe that they receive more professional recognition for the quantity of publications than for the quality of their research.

Clinical research means many things to many people. For some, it primarily means research describing the mechanisms of disease, often at the cellular level. It is often possible to conduct this kind of research well with small numbers of subjects, animals or cell systems because of the degree of control that can be exercised over studies in laboratories or clinical investigative units. In the course of explaining nature in this way, hypotheses may be raised that bear directly on clinical decisions. But it is rarely possible to test these hypotheses rigorously without resorting to studies in which a larger number of people are followed over a considerable period of time. Certainly, both kinds of research are necessary for the growth of medical knowledge. This article has described changes in the latter kind of research — research for clinicians — over the past 30 years. Although understanding of sound research designs has increased greatly in recent years, we found little evidence that this understanding has been translated into more frequent use of rigorous designs for clinical research.

REFERENCES

1. Hill AB: Controlled Clinical Trials. Springfield, Illinois, Charles C Thomas, Publisher, 1960
2. Cochrane AL: Effectiveness and Efficiency: Random reflections on health services. London, Nuffield Provincial Hospitals Trust, 1972
3. Dawber TR, Kannel WB: Coronary heart disease as an epidemiology entity. Am J Public Health 53:433-437, 1963
4. Cornfield J, Haenszel W: Some aspects of retrospective studies. J Chronic Dis 11:523-534, 1960
5. Feinstein AR: Clinical biostatistics. XLIV. A survey of the research architecture used for publications in general medical journals. Clin Pharmacol Ther 24:117-125, 1978
6. Juhl E, Christensen E, Tygstrup N: The epidemiology of the gastrointestinal randomized clinical trial. N Engl J Med 296:20-22, 1977
7. Ulrich's International Periodicals Directory. New York, RR Bowker Company, 1975
8. Feinstein AR: Clinical Biostatistics. St Louis, The CV Mosby Company, 1977
9. Fleiss JH: Statistical Methods for Rates and Proportions. New York, John Wiley & Sons, 1973, p 146
10. Dixon WJ, Massey FJ Jr: Introduction to Statistical Analysis. Third edition. New York, McGraw-Hill Book Company, 1969, p 351
11. Feinstein AR, Koss N, Austin JHM: The changing emphasis in clinical research. I. Topics under investigation: an analysis of the submitted abstracts and selected programs at the annual "Atlantic City Meetings" during 1953 through 1965. Ann Intern Med 66:396-419, 1967
12. Feinstein AR, Koss N: The changing emphasis in clinical research. III. Follow-up report for the years 1965-1969. Arch Intern Med 125:885-891, 1970
13. Kitchell JR, Glover RP, Kyle RH: Bilateral internal mammary artery ligation for angina pectoris. Am J Cardiol 1:46-50, 1958
14. Cobb LA, Thomas GI, Dillard DH, et al: An evaluation of internal-mammary-artery ligation by a double-blind technic. N Engl J Med 260:1115-1118, 1959
15. Weitzman S, Berger S: Clinical trial design in studies of corticosteroids for bacterial infections. Ann Intern Med 81:36-42, 1974
16. Gifford RH, Feinstein AR: A critique of methodology in studies of anticoagulant therapy for acute myocardial infarction. N Engl J Med 280:351-357, 1969
17. American Federation for Clinical Research: The scientific uses and abuses of the clinical trial: treatment of chronic stable angina with saphenous vein bypass grafting: Randomized Veterans Administration Cooperative Study. Clin Res 26:229-235, 1978