INFORMETRICS 89/90
L. Egghe and R. Rousseau (Editors)

A CITATION ANALYSIS OF CLINICAL TRIALS; ARE DEFINITIVE STUDIES LESS CITED THAN OTHERS?

Bluma C. PERITZ

Graduate School of Library and Archive Studies
Hebrew University, Jerusalem 91904, P.O. Box 503, Israel

Abstract

A citation analysis of 316 clinical trials published in 1979 and 1980 in two major medical journals : New England Journal of Medicine and Lancet arrives at an apparently paradoxical finding : large trials, carried out according to modern designs involving randomized and double-blind allocation procedures, tend to be less cited than other, less rigorous studies. The explanation put forward in this paper revolves around the idea that heuristic value is an *important determinant of citedness*. Since studies designed to yield final conclusions are likely to be preceded by smaller and more tentative trials, new ideas, concepts and approaches are more likely to originate in the latter rather than in the former.

1. INTRODUCTION

Although citation analysis has produced many interesting and practically useful results, the basic reason why certain papers are highly cited while others are not is still somewhat elusive. Possibly, one factor contributing to the citation strength of a scientific paper might be its heuristic value, that is, its ability to generate new concepts, ideas and hypotheses. If such were the case then papers which propose a new approach to a problem, even if based on evidence that is sketchy or preliminary, might be more cited than a large and rigorous study which says the "final word" on some specific question. Perhaps citation analysis itself might be used in order to find out to what extent such a hypothesis fits the facts. The testing ground chosen here is the area of clinical trials.

As shown in the important study by Fletcher and Fletcher [1], clinical trials are a growing field within the wider discipline of clinical research. Simply put, a clinical trial is the evaluation of the effect of some therapeutic or preventive measure, or of some other type of intervention, on the outcome of a disease in humans. As pointed out by Schwartz, Flamant and Lellouch [2], clinical trials proceed through several stages, from the first tentative use in humans after the completion of animal experiments, through the systematic comparison of the treatment in question with its alternatives, to the assessment of possible side-effects. The main issue involved in the design of a clinical trial are those of ethical acceptability and of comparability. In order to ensure comparability three aspects are of paramount importance : a) the adequate selection of a control group (in some instance the patient's pre-treatment state may serve as a control to its post-treatment state); b) the *random* (and hence unbiased) allocation of patients to treatment and control groups, and c) (if possible and ethically justifiable) the avoidance of all distortion brought about by the patient's and clinician's knowledge of the treatment administered, by means of the so-called double-blind design. In

addition, chance error in the outcomes can be kept low by increasing as far as possible the number of cases participating in the trial.

It appears, therefore, that studies that are large, adequately controlled, randomized and double-blind are best suited to strenghten our knowledge of the therapeutic measure under investigation. On the other hand, as pointed out in the previous paragraph, such studies are likely to be preceded by several smaller, more tentative studies. If heuristic value is the main determinant of "citedness" then the latter type of study is more likely to be cited by researchers; if, on the other hand, the importance of its contribution to knowledge or the thoroughness of its research procedures were more decisive then one would expect the former kind of study to be cited more often. This argument leads us, therefore, to the following, apparently paradoxical hypothesis for this paper : the more definitive a trial the less cited it is likely to be. Operationally this hypothesis translates to the statement : large, randomized and double-blind clinical trials tend to be cited less frequently than trials having neither of these characteristics.

2. METHOD

The present study includes all clinical trials - according to the definition proposed in the previous section - that were published in the *New England Journal of Medicine* or in *Lancet* in the years 1979-1980. These two publications, the first American and the second British, are both highly prestigious and much cited weekly journals. Excluded from the study population were papers *on* clinical trials (their philosophy, ethics, statistical models, etc.) which do not include any empirical data on clinical trials or do so only by way of illustration. Further exclusions were : editorials, letters to the editor (although in the latter category there were a few simple clinical trials - mainly in Lancet) and literature reviews. On the other hand, secondary analyses of a given clinical trial were included in the study.

The following variables were ascertained for each paper :

1. The main dependent variable was the number of times cited. This was retrieved on-line from *Science Citation Index*, for each year from the publication of the paper until the end of 1987. Average numbers of citations per year were then obtained, dividing by 9 for papers published in 1979 and by 8 for those appearing in 1980. As a by-product, the year in which most citations occured, the so-called peak year, was also obtained. (If the citations to a given paper attain their peak year more than once, only the latest peak year is taken into account).

2. The following independent variables were considered :
   a) whether the study was controlled or not;
   b) if controlled, whether some random allocation was carried out;
   c) whether the assignment to cases and controls was double-blind, blind (i.e. only with the patient ignoring which group he was assigned to) or non-blind;
   c) the number of cases involved in the trial was also taken as a study variable.

Several rules and decisions were necessary in order to classify the trials as accurately as possible into the above categories.

1) The group of controlled trials without randomization was taken to include 'self-controlled" trials, i.e. designs in which the final status of each patient was compared to his status before the administration of treatment. No attempt was made to assess the suitability of this (or any other) design to the trial in question. Included in this group were also the so-called

crossover designs, provided there was no random allocation.

2) Uncontrolled trials were those in which no control group was chosen. Sometimes it was difficult to decide whether a given study was uncontrolled or self-controlled; indeed in most uncontrolled studies some effort was made to study the change in the patients' status from some earlier stage to a later one. We considered a trial to be self-controlled, rather than uncontrolled, if the onset of treatment did not precede the ascertainment of the baselines data in each patient and if change in the patient's status could not be due to non-random causes other than the treatment itself. Since in several cases there could be some doubt about the latter qualification, uncontrolled and controlled non-randomized trials were pooled for the final analysis.

3) The size of the clinical trial varies from half a dozen cases to several thousands. In order to get some idea of the size of the trials these are classified into the following groups :
   a) small - up to 11 cases
   b) medium - 12 to 49 cases
   c) large - 50 to 199 cases
   d) mass - 200 or more cases.

   Typically, only the number of cases - not that of the controls - was counted. The few instances in which two treatments, rather than a treatment and a control were compared, made no difference to the above classification. Whenever three or more treatment groups are given they were all counted as cases - again with little effect on the classification. "Historical" controls, i.e. controls from previous studies, were disregarded. In the analysis small and medium studies and large and "mass" studies were combined.

4) A trial was considered randomized only if this was specifically stated, even if the context made it likely that this was the case. So-called "pseudo-random" allocation was not included, mainly because this term is not well-defined.

After the data were extracted from the two journals they were checked by another resource person who is familiar with the field of clinical trials, and reviewed with the author.

3. STATISTICAL ANALYSIS

The results of this investigation are presented in the usual manner, through average numbers of citations per year and the corresponding standard deviations. However, since citation distributions are known to be very skewed, a further non parametric analysis was required.

Following Peritz [3], weighted averages of Mann-Whitney statistics were used. The Mann-Whitney statistic U for the comparison of two samples is defined as the number of times an item (a paper) in the first sample has a higher value (a higher number of citations) than a paper from the second sample (Siegel [4]). An indication of the extent to which the first sample is "higher" than the second sample is given by :

$$\frac{(2U - mn)}{mn}$$

where m and n are the sizes of the first and second sample, respectively. This indicator varies between -1 and 1 and takes on the value zero when the two samples are distributed identically. An approximately normal statistic based on U is :

$$z = \frac{U - mn/2}{\sqrt{mn(m+n+1)/12}}$$

In this study one needs to average several U-values, one for each journal year and perhaps some other characteristic of the paper. An optimal weighted average of k such subgroups or states is :

$$\frac{\Sigma_i (U_i - m_i n_i/2)(m_i + n_i + 1)^{-1}}{\Sigma_i\ m_i n_i (m_i + n_i + 1)^{-1}} \qquad i = 1,\ldots,k$$

and the z-statistic to be used for hypothesis testing is, according to Van Elteren [5] :

$$z = \frac{\Sigma_i (U_i - m_i n_i/2)(m_i + n_i + 1)^{-1}}{(\Sigma_i\ m_i n_i/(m_i + n_i + 1))^{1/2}} \sqrt{12}$$

4. RESULTS

The numbers of papers on clinical trials, classified according to their characteristics are presented in Table 1.

Table 1 : Papers on clinical trials by journal, year and characteristics of the paper

Type of Study	New England J. of Med.		L a n c e t		Total
	1979	1980	1979	1980	
Uncontrolled					
Small/Medium	9	11	24	25	69
Large/Mass	2	4	4	9	19
Total	11	15	28	34	88
Controlled non-randomized					
Small/Medium	11	18	24	27	80
Large/Mass	5	4	10	17	36
Total	16	22	34	44	116
Randomized					
Small/Medium	5	3	14	22	44
Large/Mass	12	16	15	25	68
Total	17	19	29	47	112
Therefrom : Double-blind*	11	10	15	24	60
Grand total	44	56	91	125	316

• In addition, there were 19 double-blind studies in which the allocation to treatment groups was not specifically stated to be random.

It is readily seen that more clinical trials are published in *Lancet* than in the *New England Journal of Medicine*. In both journals there is a definite increase in the number of clinical trials from 1979 to 1980. The papers are about equally divided between "uncontrolled, non-randomized" and "randomized" studies; this holds for each of the two journals and the two years of publication. Over one half of the randomized clinical trials are carried out double-blind; here too there is little variation between journals and years. There are also 19 papers in which allocation is double-blind and yet there is no mention of randomization. The proportion of small to medium trials decreases from close to 80 % among the "uncontrolled" studies through the "controlled, non-randomized" groups to about 40 % in the randomized studies.

Table 2 summarizes the total numbers of citations included in the study, for each journal and year of publication of the source paper. Taking into account the fact that for papers published in 1979 there were about nine years of exposure to citation while the corresponding number of years for 1980 was eight, the average number of citations per paper and per year since publication was 8.0.

Table 2 : Source papers and citations by journal and year of publication

Journal	Total number of papers	Total number of citations
New England Journal of Medicine, 1979	44	4875
New England Journal of Medicine, 1980	56	4944
Total	100	9819
Lancet, 1979	91	5712
Lancet, 1980	125	5892
Total	216	11604
Grand total	316	21423

The distribution of the peak years of citation is given in Table 3. In almost two thirds of the papers the peak year was between three and five years after publication.

Table 3 : Number of papers by peak citation year and by year of publication of source paper*

Year of Source paper	Peak Citation Year									
	1979	1980	1981	1982	1983	1984	1985	1986	1987	Total
1979	2	7	26	35	17	9	22	10	-	128
1980	-	2	13	48	32	25	21	17	10	168
Percentages										
1979	1.6	5.5	20.3	27.3	13.3	7.0	17.2	7.8	-	100.0
1980	-	1.2	7.7	28.6	19.0	14.9	12.5	10.1	6.0	100.0

* Number of uncited papers : 7 in 1979 and 13 in 1980.

As shown in Table 4, the average number of citations per paper and per year is definitely lower for *Lancet* than for the *New England Journal of Medicine*. This is the case for each trial size and for each type of study : randomized, controlled, unrandomized, or uncontrolled. In *Lancet* the small to medium sized trials tend to be more cited than the larger ones; this is, however, not necessarily the case for the *New England Journal of Medicine*. Throughout this study the standard deviations tend to be of the same order of size as the averages; this is indicative of the skewness of the citation distributions and in particular of the presence of outliers.

Table 4 : Average number of citations by year, by journal, year of source paper and characteristics of the paper*

Type of Study	New England J. of Med.		Lancet		N.E.J. Med.	Lancet
	1979	1980	1979	1980	Total	Total
Uncontrolled						
Small/Medium	11.1(8.4)	7.6(6.2)	8.8(7.5)	7.2(9.1)	9.2	8.0
Large/Mass	12.1(5.7)	14.4(13.3)	5.4(4.6)	3.4(2.4)	13.7	6.4
Controlled non-randomized						
Small/Medium	14.0(8.4)	11.2(7.0)	7.4(7.5)	4.9(3.9)	12.3	6.1
Large/Mass	23.1(14.0)	18.6(14.8)	4.9(4.9)	5.6(6.6)	21.1	5.3
Randomized						
Small/Medium	14.6(6.6)	9.5(2.1)	5.6(6.7)	6.4(8.3)	12.7	6.1
Large/Mass	6.1(6.1)	10.6(8.8)	6.1(6.6)	5.9(5.5)	8.7	6.0
Therefrom : Double-blind*	8.3(7.3)	8.8(7.5)	7.5(6.7)	4.2(3.7)	5.7	5.5
Grand Total	12.3	11.0	7.0	5.9	11.6	6.4

* Standard deviations in parentheses.

Table 5 presents means and standard deviations for the papers which form the "target group" of this study : the clinical trials which are large or "mass", randomized *and* double-blind. It is seen that, with the exception of Lancet 1979 the averages are lower than the corresponding values in Table 4.

Table 5 : Average number of citations in all clinical trials that are, large or "mass", randomized *and* double-blind, by journal and year of source paper

Journal	Total number of papers	Average	S.D.
New England Journal of Medicine, 1979	7	5.7	6.3
New England Journal of Medicine, 1980	9	9.1	8.1
Lancet, 1979	8	7.9	7.4
Lancet, 1980	11	3.4	3.0

As pointed out in the previous section, the comparisons suggested by our hypotheses are carried out by nonparametric methods : the average Mann-Whitney statistics of [5] above and the corresponding z-statistic and (one-sided) P-value, controlling for journal and year of publication of the trial. The results are given in Table 6.

Table 6 : Average Mann-Whitney statistics and tests for various comparisons*

Type of Study	M.-W. statistics	z	P
1. Randomized, double-blind, large or mass versus all other papers	0.18	1.747	0.04
2. Randomized, double-blind, large or mass versus non randomized, non-blind, small or medium	0.22	1.916	0.03
3. Randomized versus non-randomized controlling for number of cases	0.09	1.155	0.12
4. Large or mass versus small or medium controlling for randomization	0.05	0.746	0.22
5. Double-blind versus other, controlling for number of cases (randomized only)	0.26	2.285	0.01

* For formulas see section 3.

The comparison of our target group of clinical trials to all the other studies included here yielded a Mann-Whitney statistic of 0.18; the corresponding test was significant at the 5 % level. If the target group is compared to all the trials that have the converse characteristics : *non-randomized, non-blind and* small or medium, the Mann-Whitney statistic is 0.22 and the P-value of the corresponding test is 0.03.

An attempt was made to isolate the effect of each of the target group's characteristics : randomization, size (small-medium versus large-mass) and blindness, controlling for one or two of the other variables. The comparison of randomized with non randomized trials controlling for size (as well as for the journal and year of publication) yielded a Mann-Whitney statistic of 0.09 and a z-value that was not significant, but still pointing in the direction of the alternative. The same is true for the comparison between large or "mass" trials and small or medium trials, controlling for randomization. Finally, comparing double-blind with non-double-blind trials, controlling for size and confining oneself to studies with randomization, gave a Mann-Whitney statistic of 0.26; the corresponding test was significant with a P-value of 0.011. The nineteen cases which were double-blind but for which it was not specifically mentioned that the allocation was random are not included in this calculation, since one cannot be certain that randomization was not taken for granted and merely omitted from the presentation. In any case, the inclusion of these cases would not affect the results materially.

5. DISCUSSION

The findings of the previous section suggest that large, randomized, double-blind clinical trials are significantly less cited than other trials. Since the former are usually further ahead on the road to a final conclusion

regarding the therapeutic measure in question than the latter, one wonders why more definitive studies should be less cited than more tentative ones. The interpretation offered in the introduction to this paper was that tentative and preliminary studies are heuristically more important - and hence more citable - than studies which by their very conclusiveness are apt to "wrap-up" some research issue. Moreover, small and tentative investigations are more likely to inspire clinical researchers with limited time and resources to follow up on the ideas proposed there.

No doubt, other interpretations of this finding are possible. Thus, one might speculate that the truly "burning" issues regarding therapeutic measures often cannot be randomized or double-blind, for ethical reasons. However, if this were so, one would still have to postulate that trials dealing with "burning" issues are cited more often than others. Another possible supposition is that interesting medical issues often preclude large-scales trials because of the rarity of the disease involved. This may well be true; however, it was shown in Table 6 that randomization and blindness have separate effects (not both of them statistically significant) after controlling for size. One is led, therefore to accept, at least tentatively, the idea that heuristic value is a determinant of a paper's citedness. More precise conclusions could be reached only through a direct investigation of the role of citations in various scientific fields.

One final word on choosing clinical trials for the present study. The main reason was that clinical trials are essentially an applied field in which the prevalent attitude is one of solving specific, clearly defined problems. It is not clear whether a study carried out in some area of basic science, in which a typical paper raises as many questions as it solves, would yield similar results.

ACKNOWLEDGEMENTS

I would like to thank Mrs. Edna Granit, Dr. Yecheskel Wolman and an unnamed epidemiologist for their valuable help in carrying out this research.

REFERENCES

[1] Fletcher, R.H., Fletcher, S.W., Clinical research in general medical journals; a 30-year perspective. New England Journal of Medicine, 301(4), 180-183, 1979.

[2] Schwartz, D., Flamant, R., Lellouch, J., Clinical trials, tr. by M.J.R. Healy, London : Academic Press, 1980.

[3] Peritz, B.C., Are methodological papers more cited than theoretical or empirical ones? The case of sociology. Scientometrics, 5(4) : 211-218, 1983.

[4] Siegel, S., Nonparametric statistics for the behavioral sciences. New York : McGraw-Hill, 1956.

[5] Van Elteren, P., On the combination of independent two sample test of Wilcoxon. Bulletin de l'Institut International de Statistique, 37 : 351-361, 1960.