Chula Med J Vol. 45 No. 4 April 2001

Sequential analysis : comparison of two proportions

วีนัส อุดมประเสริฐกุล *

Udomprasertgul V. Sequential analysis : comparison of two proportions. Chula Med J 2001 Apr; 45(4): 367 - 75

In the usual type of experimental study, particularly in clinical research, the number of subjects to be included in the various treatment groups under comparison is often determined by the availability of these subjects, funds and the intervention being studied. However, ethical considerations demand that a trial be stopped as soon as there is clear evidence that one of the treatments is superior, if the results do not look promising or if an impressive difference is already apparent, This requires sequential analysis, where feasible.

This section covers only sequential plans of binary preference and comparison of two proportions. This includes how to calculate the boundary lines for a one - sided sequential plan and a two-sided sequential plan, and how to consider the different advantages and disadvantages of sequential testing.

Key words : *Sequential analysis, Binary, Preference, Proportion.*

Reprint request : Udomprasertgul V. Institute of Health Research, Chulalongkorn University, Bangkok 10330, Thailand.

Received for publication. January 27, 2001.

* Institute of Health Research, Chulalongkorn University

วีนัส อุดมประเสริฐกุล. การวิเคราะห์ข้อมูลตามลำดับ : แบบการเปรียบเทียบอัตราส่วน. จุฬาลงกรณ์เวชสาร 2544 เม.ย; 45(4): 367 - 75

ตามปกติทั่วไปการวิจัยเชิงทดลองโดยเฉพาะทางคลินิก จำนวนขนาดตัวอย่างที่ต้องการในแต่ละกลุ่มทดลอง จะต้องพิจารณาถึงความเป็นไปได้ของตัวอย่างที่ศึกษา และงบประมาณ อย่างไรก็ตามควรคำนึงถึงในด้านจริยธรรมทางการแพทย์ด้วยว่า การทดลองใด ๆ ควรจะหยุดทันที ถ้ามีข้อบ่งชี้ที่ชัดเจนว่า การทดลองชนิดนั้นได้ผล หรือผลการศึกษาไม่เป็นไปตามที่คาดหวัง วิธีวิเคราะห์ที่ต้องการทราบผลรวดเร็วคือการวิเคราะห์ข้อมูลตามลำดับ

ในบทความนี้จะครอบคลุมเฉพาะการวิเคราะห์ข้อมูลตามลำดับที่ตัววัดเป็นแบบ 2 คำตอบ และเปรียบเทียบอัตราส่วนเท่านั้น ซึ่งรวมทั้งการคำนวณเส้นขอบเขตของการวิเคราะห์ข้อมูลตามลำดับแบบข้างเดียวและสองข้าง ตลอดจนข้อดีและข้อเสียของการวิเคราะห์ข้อมูลตามลำดับ

Sequential analysis is a technique wherein one conducts a statistical test of significance sequentially over time as the outcome data are collected. After each observed outcome, one analyzes the cumulative data and reaches one of the following three decisions:

(1) stop the data collection, reject the null hypothesis, and claim that the active treatment is efficacious;

(2) stop the data collection, do not reject the null hypothesis, and claim that the active treatment is not superior;

(3) continue the data collection, as the cumulative data are inadequate to draw a firm conclusion.

In the usual type of experimental study, particularly of a clinical nature, the number of subjects to be included in the various treatment groups under comparison is often determined by the availability of these subjects, funds, or the intervention being studied. More correctly, however, the size of the groups used should depend on the anticipated difference in treatment effects, or on the size of the treatment difference considered important by the investigator. The size of the groups will also depend on the magnitude of the probability of error in rejecting or accepting the working null hypothesis.

In many experiments, the investigator would like to check the results before the preassigned total number of subjects has been studied. Thus one could perform one or more significance tests part way through the study, and make a decision about the need for, and the feasibility of, continuing the experiment. The study might then terminate if the results did not look promising or an impressive difference was already apparent. Unfortunately, multiple testing changes the required probabilities of error. A desire to check the results periodically and still control the probabilities of error leads logically to a sequential type of design.

During the Second World War the need for determining as rapidly as possible which of two products was better gave impetus to the development of the theory and application of sequential analysis by Wald and his co-workers (1947-1954). Since that time, a number of papers dealing with the applications of these methods in medical experimentation have appeared.

Armitage (1960) has been concerned with the design of experiments for the comparision of alternative medical treatments. He argues that ethical considerations demand that a trial be stopped as soon as there is clear evidence that one of the treatments is to be preferred, and this requires sequential analysis, where feasible.

Sequential Plans of Binary Preferences

In many clinical trials it is appropriate to assess the relative merits of two treatments by collecting a series of qualitative preferences in favour of one or other treatment. If two treatments, A and B, are equally effective, a suitable experimental design with randomization will ensure that the observed series of preferences is a random binary sequence, in which the probability of a preference in favour of A is 1/2. By a 'random binary sequence' we mean a series in which the A's and B's alternate with no systematic pattern. In situations where the preferences form a binary sequence, we shall denote by θ the probability that a preference is in favour of A; that is, in an indefinitely long sequence, θ would be the proportion

of A preferences. The null hypothesis, then, is that $\theta = 1/2$.

If A is really better than B, so that the null hypothesis is not true, the proportion of A preferences in the long run will be greater than 1/2; and if B is preferable, the proportion of A preferences will be less than 1/2. It may be, under these circumstances, that the preferences no longer form a binary sequence with a constant value of θ. For instance, if the type of patient entering the trial changes gradually as the trial proceeds, θ may be higher (or lower) at the beginning of the trial than at the end. Nevertheless, we can consider, as one possible departure from the null hypothesis, a situation in which θ takes some constant value, say θ_1 , greater than 1/2; or if B is the better treatment, a constant value, θ_0, less than 1/2.

We could then require that our sequential procedure has a specified overall significance level, 2α, and has a high power of detecting a change of θ from 1/2 to θ_1, or from 1/2 to θ_0.

For simplicity we shall suppose that θ_0 and θ_1 are symmetrical about the value1/2; that is, $\theta_0 = 1 - \theta_1$.

For the case of preferences in favour of A or B derived from paired observations, the application of the sequential test procedure is in the form of a 'chart'. The vertical axis measures the current imbalance in preferences in favour of each treatment. This quantity is plotted as each pair of results become available. Pairing here could be via a cross-over design in which a single patient tested in a random order on treatment A and B, or previously matched pairs of patients randomly allocated to A or B within pairs. The imbalance in favour of A or B is plotted against the number of pairs tested to form the sequential chart. Pre-assigned decision barriers are built into the chart to achieve the required error probabilities over the life of the trial. When the imbalance count crosses a barrier the associated conclusion results, either A is better than B, B is better than A, or, in some schemes, that A and B are equivalent.

Sequential Plan for Comparison of Two Proportions

In many clinical trials the crucial assessment is made by comparing two proportions: for example, the proportion of patients who show a certain degree of improvement with drug A and a corresponding proportion of improvement with drug B. To apply the sequential procedures described earlier we require that observations be made in pairs. For within-subject comparisons, a natural method is to pair two successive observations on the same subject, the order of allocation to the two treatments being determined at random. For between-subject comparisons, successive subjects entered into the trial can form a pair, the allocation to the two treatments within such a pair being again at random. The allocation for each pair of observations can be done by tossing a coin. Alternatively, a table of random sampling numbers can be used. If the two treatments are denoted by A and B, an even random digit can be taken to indicate the order AB and an odd digit indicating BA. Thus one random digit is required for each pair of observations. In the sequential analysis, the first A in each balanced group is paired with the first B; the second A with the second B; and so on. The results for each pair will fall into one of the four following categories (S denoting success, F denoting failure) :

Type	Treatment		Preference	
	A	B		
(a)	S	S	-	Tied pairs
(b)	F	F	-	
(c)	S	F	A	Untied pairs
(d)	F	S	B	

Pairs of types (a) and (b) are called tied pairs, because the comparison between treatments results in a tie; those of types (c) and (d) may similarly be called untied pairs. The untied pairs result in a preference for one or other treatment: type (c) a preference for A, and type (d) a preference for B.

Suppose that the probability of success with treatment A is π_1 and the probability of success with B is π_2.

Then, on the assumption that pairing is random, we can write down the probability of each of the four types of pair:

	Type	A	B	Preference	Probability
Tied	(a)	S	S	-	$\pi_1 \pi_2$
	(b)	F	F	-	$(1 - \pi_1)(1 - \pi_2)$
Untied	(c)	S	F	A	$\pi_1 (1 - \pi_2)$
	(d)	F	S	B	$(1 - \pi_1) \pi_2$

The probability of an A preference is the same as the probability that an untied pair is of type (c), namely

$$\theta = \frac{\pi_1 (1 - \pi_2)}{\pi_1 (1 - \pi_2) + (1 - \pi_1) \pi_2}$$

If $\pi_1 = \pi_2$ (the null hypothesis), $\theta = 1/2$; if $\pi_1 > \pi_2$, $\theta > 1/2$ and if $\pi_1 < \pi_2$, $\theta < 1/2$.

The probability that a pair is untied and, therefore, provides a preference, is $\pi_1 (1 - \pi_2) + (1 - \pi_1) \pi_2$.

Table 1. gives the values of θ corresponding to various combinations of π_1 and π_2.

The main points to notice are:

(a) when $\pi_1 = \pi_2$, $\theta = 1/2$;

(b) when π_1 is greater than π_2, $\theta > 1/2$;

(c) when π_1 is less than π_2, $\theta < 1/2$;

(d) for a given value of θ, the absolute difference between the probabilities of success, $\pi_1 - \pi_2$ or $\pi_2 - \pi_1$, is highest when the average of π_1 and π_2 is 1/2, and decreases as this average approaches zero or unity. Another way, is to consider a given value of the difference $\pi_1 - \pi_2$ (supposing π_1 to be greater than π_2) and use Table 1 to determine θ.

For given values of π_1 and π_2, the formulae and tables will give a value of θ, and then boundaries can be drawn depending upon which plan is chosen. If these tables are not available, boundary lines can be calculated.

The parallel lines constituting the boundaries of the one-sided sequential plan are calculated from the formulae below.

The following parameters are calculated:

$$a = \log \frac{(1-\beta)}{\alpha}$$

$$b = \log \frac{(1-\alpha)}{\beta}$$

$$c = \log \frac{1}{2(1-\theta)}$$

$$d = \log \frac{\theta}{1-\theta}$$

Where θ is the proportion of untied pairs favourable to preparation A.

i.e. $$\theta = \frac{\pi_1 (1 - \pi_2)}{\pi_1 (1 - \pi_2) + (1 - \pi_1) \pi_2}$$

We then calculate the boundary lines for the sequential plan which must be crossed before a decision can be reached:

Reject $\theta = 1/2$ and declare preparation A is better than B when

$y > a/d + (c/d)n$, ... upper line

where y is the number of untied pairs favouring the new preparation A and n is the number of untied pairs.

Reject $\theta = \theta_1$ and declare preparation A not better than B when

$y < -b/d + (c/d)n$, ... lower line

These two lines are drawn on squared arithmetic graph paper. The abcissa is n, the number of untied pairs, and the ordinate is y, the number of these pairs favourable to the new preparation A; see the example in Figure 1 and conclude that preparation A is better than preparation B.

In the previous sequential plan for comparing the effectiveness of two preparations, we were interested only in determining whether preparation A was better than preparation B and not whether B was better than A. This one-sided test is frequently encountered in clinical research when a new treatment is to be compared with an established treatment. Sometimes, however, the investigator has two competing treatments or preparations, and he wants to know which one is the better. This latter situation requires the setting-up of a two-sided plan.

The simplest two-sided sequential plan is obtained by combining two one-sided sequential plans: Plan 1, to test whether A is better than B as previously illustrated, and Plan 2, to test whether A is worse than B (this latter is equivalent to testing whether B is better than A). If the investigator selects $\alpha = 0.05$ for the

Table 1. Values for θ for Various Combination of π_1 and π_2

$\pi_1 - \pi_2 = 0.10$			$\pi_1 - \pi_2 = 0.20$			$\pi_1 - \pi_2 = 0.30$		
π_1	π_2	θ	π_1	π_2	θ	π_1	π_2	θ
0.11	0.01	0.924	0.21	0.01	0.963	0.31	0.01	0.978
0.15	0.05	0.770	0.25	0.05	0.864	0.35	0.05	0.911
0.20	0.10	0.692	0.30	0.10	0.794	0.40	0.10	0.857
0.35	0.25	0.618	0.45	0.25	0.710	0.50	0.20	0.800
0.55	0.45	0.599	0.60	0.40	0.692	0.65	0.35	0.775
0.75	0.65	0.618	0.75	0.55	0.710	0.80	0.50	0.800
0.90	0.80	0.692	0.90	0.70	0.794	0.90	0.60	0.857
0.95	0.85	0.770	0.95	0.75	0.864	0.95	0.65	0.911
0.99	0.89	0.924	0.99	0.79	0.963	0.99	0.69	0.978

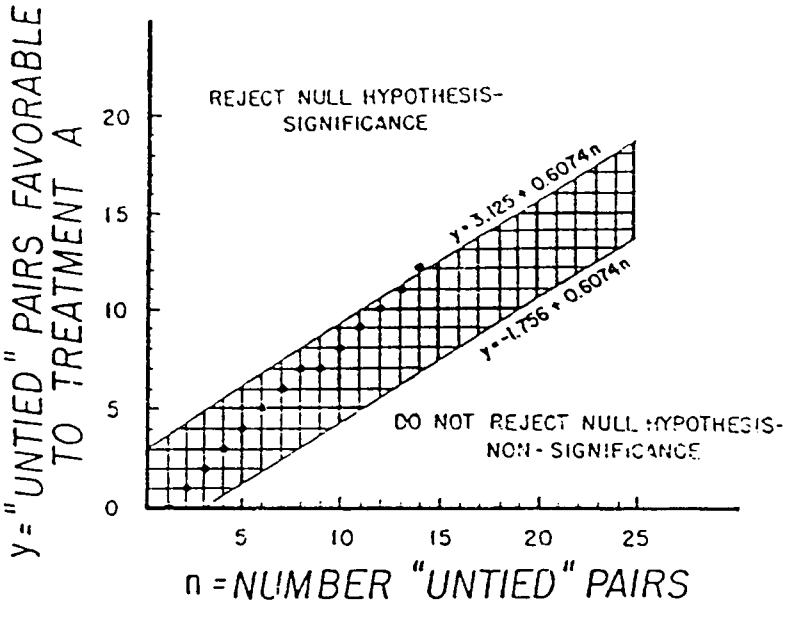

Figure 1. One-sided sequential plan, $\alpha = 0.05$; $\beta = 0.20$; $\theta = 0.7083$

one-sided plan, then it will be necessary to apportion half of the α to each side of the plan, so that the new α for each of the one-sided plans to be constructed will be 0.025.

Plan 1 will provide the two lines for the upper portion of two-sided plan. We calculate values for a, b, c, and d from the same formulae as the one-sided plan.

The lines for the upper portion of the two-sided sequential plan are:

$y_1 = a/d + (c/d)n$, ... upper line

and

$y_0 = -b/d + (c/d)n$, ...lower line

where y is the number of untied pairs favouring preparation A and n is the number of untied pairs.

Plan 2 will provide the two lines for the lower portion of the two-sided plan. We obtain the same values for a, b, c, and d as for Plan 1.

The lower pair of parallel boundaries corresponding to Plan 2 are:

$y_0' = b/d + (1 - c/d)n$, ... upper line

and

$y_1' = -a/d + (1 - c/d)n$, ... lower line

These two pairs of parallel lines are drawn on graph paper, as shown in Figure 2.

Advantages and Disadvantages of Sequential Tests

Sequential statistical testing is not commonly used in applied research. While the conceptual advantates put forward by Armitage and others are persuasive, there are also disadvantages with the approach. Some statisticians are even opposed on theoretical/philosophical grounds (Anscombe F.J, 1963) but they tend to be in the minority. In the section below we will look at the advantages and disadvantages of sequential testing.

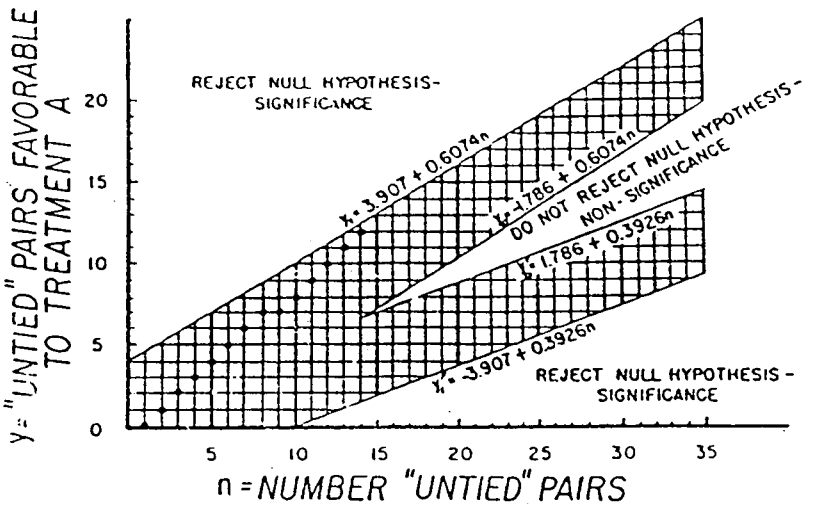

Figure 2. Two-sided sequential plan, $\alpha = 0.025$; $\beta = 0.20$; $\theta = 0.7083$

Advantages

There are three basic advantages of the sequential method.

(1) The sample size necessary to assert significance is less, on average, than the corresponding sample size calculated on the basis of a non-sequential design.

(2) If the difference between two preparations, A and B, exceeds that assumed in the preparation of the plan or if, in fact, A is worse than B instead of being better than B as specified in the alternative hypothesis, then a decision of non-significance is reached more rapidly.

(3) Sequential methods are, therefore, particularly suitable for trials to asses the treatment of acute conditions, or those in which rapid relief of a chronic condition is sought.

Disadvantages

(1) One of the basic disadvantages is that the nature of the study may not lend itself readily to this method, as, for example, when there is an appreciable time lag between the intake of the subjects and the evaluation of the results, so that many subjects may still be in the process of evaluation when a decision is reached.

(2) Another problem which is sometimes encountered is that, because of random variation, the testing may need to go on much longer than the calculated average sample size. To circumvent this difficulty, the investigator can decide in advance to stop the experiment at, for example, twice the calculated fixed sample size and make the decision corresponding to the nearest line. If this truncation is performed far out on the sequential graph, such as at twice the fixed sample size, it does not materially change the values previously selected for α and β.

(3) Trials of long-term treatment of chronic disease, where the measurement of response can be made only after a long follow-up period, are much less amenable to sequential methods.

(4) While sequential plans have been worked

out for a fairly large number of situations, they have not been satisfactorily worked out for some of the more complicated designs that one engages in, such as factorial designs. Sequential analysis is still pretty much confined to simple types of design.

(5) In large-scale trials involving the cooperation of many centres, a sequential analysis may present some organizational problems; it may be a little more troublesome to arrange for records to be sent to some central point, throughout the trial and without undue delay, than to leave the collection of records until the end.

(6) One further disadvantage of the sequential method is largely an administrative one. In setting up a study it is often necessary to prepare a suitable budget and also to hire additional personnel specifically for the study. These people are usually hired for a definite period of time. If the study is terminated more rapidly than originally planned, or if it continues longer than originally planned, the problem arises as to what to do with the staff.

References

1. Anscombe FJ. Sequential medical trials. J Am Statist Assoc 1963; 58: 365 - 83
2. Armitage P. Statistical Methods in Medical Research. Oxford: Blackwell, 1971: 415 - 25
3. Armitage P. Sequential Medical Trials. 2nd ed. New York: John Wiley and Sons, 1975: 1 - 194
4. Bross IDJ. Sequential medical plans. Biometrics 1952; 8(3): 188 - 205
5. Bross IDJ. Sequential clinical trials. J Chron Dis 1958; 8(3): 349 - 65
6. Chilton NW, Fertig JW, Kutscher AH. Studies in the design and analysis of dental experiments. III. Sequential analysis (double dichotomy). J Dent Res 1961; 40(2): 331 - 40
7. Chilton NW. Design and Analysis in Dental and Oral Research. Philadelphia: J.B. Lippincott , 1967: 260 - 84
8. Colton T. Statistics in Medicine. Boston: Little, Brown,1974: 229 - 36
9. Fertig JW. Clinical trials: Fixed sample size; sequential analysis, adaptive design. Strat Clin Res 1976; 9: 26 - 36
10. Fraunhofer JA, Murray JJ. Statistics in Medical, Dental and Biological Studies. London: Tri-Med Books, 1976: 87 - 90
11. Wald A. Sequential Analysis. New York: John Wiley and Sons, 1947: 1 - 174
12. Wetherill BG. Sequential Methods in Statistics. New York: John Wiley and Sons, 1975: 1 - 232