Thorax 1986;**41**:830–832

Number of patients required in lung function studies

ANTHONY G DAVISON, P M FAYERS, A J NUNN, K M VENABLES, A J NEWMAN TAYLOR

From the Cardiothoracic Institute and the Medical Research Council Tuberculosis and Chest Diseases Unit, Brompton Hospital, London

ABSTRACT Tables are presented showing estimates of the number of subjects which is required to give an 80% or 90% chance of detecting various differences in forced expiratory volume in one second, forced vital capacity, total lung capacity, transfer factor, and residual volume between the mean of two groups by means of Student's *t* test.

There is increasing concern about the interpretation of "negative" investigations.[1-3] Important differences may not be detected because an inadequate number of subjects is studied. The number required should be considered carefully in the planning stages of a study to avoid this pitfall. We have constructed tables that may be used to estimate the number of subjects required in studies of lung function. These are useful in comparisons of the FEV_1, forced vital capacity (FVC), total lung capacity (TLC), transfer factor (TLCO) or residual volume (RV) in two independent groups with Student's unpaired *t* test.

The number of subjects required depends on the following five interrelated factors.

The level of significance selected In comparisons of independent groups with Student's *t* test the null hypothesis is that no difference exists. If an observed difference is "significant at the 5% level" for example, on the null hypothesis there is a probability of 5% or less that the observed difference could be due to chance; only 5% or fewer repeat studies would obtain such a large difference by chance. Obtaining a result that is significant at a given level where the difference is due to chance is termed the type I error (α). It is not the purpose of this note to give a formal definition of p values and significance tests; details of such terms may be found in standard statistical texts.[4]

The power of the test In some studies a real difference is not detected because the observed difference between the means fails to attain the 5% significance level. This is termed a type II error (β), and usually arises because the study population is too small to detect an important difference. A 20% chance of a type II error means that there is a 20% chance of failing to find a difference at the significance level selected when in fact a true difference exists. This is usually expressed in terms of the power of the test $(1-\beta) \times 100\%$. Thus a power of 80% means that there is an 80% chance of detecting a difference at a given significance level if a real difference exists.

If the significance level is reduced—for example, from the 5% level to the 1% level—the number of patients needed will increase if the power is to remain the same. Powers of 80–90% are usually selected as the minimum acceptable.

The standard deviation of the measurement It is important to obtain an estimate of the standard deviation and this may be made from previous studies or a pilot study. The standard deviations used in the tables in this paper are derived from a pooled series of published normal values.[5] Since normal values of lung function are highly dependent on the subjects' height and age, these standard deviations apply to differences from "expected" values. Any analyses of the data should take age and height into account.

The size of the difference between the means The smaller the difference between the means the larger is the number of subjects needed to detect it. A very small true difference can always be shown to be significant given suffcient numbers. The investigator must decide what degree of difference is clinically important in the study.

The relative numbers of subjects in the groups The total number of subjects required is least when the groups are of equal size. This is assumed in the tables. In clinical trials equal sized groups are often used, to minimise the total sample size required. In epidemiological studies, however, this does not necessarily apply: for example, it may be easier to obtain subjects from one population than from another. In such cases a generalised form of the equation should be used.[6]

Address for reprint requests: Dr Anthony G Davison, Department of Occupational Medicine, Cardiothoracic Institute, Brompton Hospital, London SW3 6HP.

Accepted 23 June 1986

Calculation of number of patients

The usual approximation used for estimating the number of patients required for a study is given below[4 6]; this equation applies to the comparison of two equal sized groups when the unpaired t test is used. A more precise method is available, using the non-central t distribution. The difference in the estimations of numbers of patients required is small. The equation below can be readily solved by the investigator; the solution of the equation using the non-central t distribution is considerably more difficult.

The equation for calculating the number of patients is:

$$2N = \frac{4\sigma^2(z\alpha + z\beta)^2}{d^2},$$

where $2N$ = total number of subjects; σ = standard deviation; $z\alpha$ = the normal deviate for the significance level ($z\alpha = 1{\cdot}96$ with $\alpha = 0{\cdot}05$ and a two sided test and $z\alpha = 2{\cdot}58$ with $\alpha = 0{\cdot}01$); $z\beta$ = the normal deviate for the power ($z\beta = 1{\cdot}28$ with 90% power and 0·84 with 80% power); d = specified difference between means.

Use of the tables

The first column shows the total number of subjects required. The number in each group will be half this. The difference in means that can be detected at the 5% level of significance, if such a difference exists, with 90% and 80% powers is shown in the next two columns. There are separate tables for men and women because of their different standard deviations. The tables can be used in two ways. Firstly, they can be used in estimating how many subjects will be required for ensuring an 80% or 90% chance of detecting a given difference in the lung function measurement if such a difference exists. Secondly, given a population of a certain size, they indicate how small a difference in lung function can be detected with 80% or 90% power. The use of the tables is illustrated by the following examples.

EXAMPLE 1

A study is planned to compare the mean FEV_1 values of a group of male asbestos workers and of a control group. It is thought that the minimum important difference would be about 200 ml. At least 200 subjects, 100 in each group, would be required to have an 80% chance of detecting this difference at the 5% level (table 1). The total number of subjects rises to about 300 if the chance of detecting the difference is raised to 90%.

Table 1 *Differences in FEV_1 detectable at the 5% significance level, with 90% and with 80% power, for given numbers of subjects*

	Difference in FEV_1 (ml)			
	Men		*Women*	
Total number of subjects (2N)	*90%*	*80%*	*90%*	*80%*
1000	105	91	78	68
750	121	105	90	78
500	148	128	111	96
400	166	143	124	107
300	191	165	143	123
200	234	203	175	151
150	270	234	202	174
100	331	286	247	213
90	349	302	260	225
80	370	320	276	239
70	396	342	295	255
60	427	369	319	275
50	468	405	349	302
40	523	452	390	337
30	604	522	450	389
20	740	639	551	477
	SD = 510		SD = 380	

EXAMPLE 2

An investigator wishes to compare the mean TLC of male manual and office workers. He has immediate access to 15 manual and 15 office workers. Table 3 shows that the true difference would have to be 829 ml or more for him to have a 90% chance of detecting a difference. The investigator considers that the difference is likely to be less than this but that he could expect a difference of at least 400 ml. He decides to postpone his study until he has 75 manual and 75 office workers to study.

Table 2 *Differences in forced vital capacity (FVC) detectable at the 5% significance level, with 90% and with 80% power, for given numbers of subjects*

	Difference in FVC (ml)			
	Men		*Women*	
Total number of subjects (2N)	*90%*	*80%*	*90%*	*80%*
1000	126	109	89	77
750	145	125	102	88
500	177	153	125	108
400	198	171	140	121
300	229	198	161	140
200	280	242	198	171
150	323	280	228	197
100	396	342	279	241
90	417	361	294	254
80	443	383	312	270
70	473	409	334	288
60	511	442	360	312
50	560	484	395	341
40	626	541	441	381
30	723	625	509	440
20	885	765	624	539
	SD = 610		SD = 430	

Table 3 *Differences in total lung capacity (TLC) detectable at the 5% significance level, with 90% and with 80% power, for given numbers of subjects*

	Difference in TLC (ml)			
	Men		Women	
Total number of subjects (2N)	90%	80%	90%	80%
1000	144	125	124	107
750	166	144	143	123
500	203	176	174	151
400	227	197	195	169
300	263	227	225	195
200	321	278	276	238
150	371	321	318	275
100	454	393	389	337
90	479	414	411	355
80	508	439	435	376
70	543	469	465	402
60	586	507	503	435
50	642	555	551	476
40	718	621	616	532
30	829	717	711	614
20	1015	878	870	752
	SD = 700		SD = 600	

Table 4 *Differences in residual volume (RV) detectable at the 5% significance level, with 90% and with 80% power, for given numbers of subjects*

	Difference in RV (ml)			
	Men		Women	
Total number of subjects (2N)	90%	80%	90%	80%
1000	85	73	72	63
750	98	84	83	72
500	119	103	102	88
400	133	115	114	99
300	154	133	132	114
200	188	163	161	139
150	218	188	186	161
100	266	230	227	197
90	281	243	240	207
80	298	257	254	220
70	318	275	272	235
60	344	297	293	254
50	376	325	321	278
40	421	364	359	311
30	486	420	415	359
20	595	514	508	439
	SD = 410		SD = 350	

Comment

The tables give estimates of the number of subjects required to obtain a statistically significant result but the accuracy of this estimate is dependent on how closely the estimated standard deviation is related to the true standard deviation in the study. We have used an estimate of the standard deviation derived from reference populations; the standard deviation of the measurements in those with respiratory disease may be greater.

Table 5 *Differences in carbon monoxide transfer factor (TLCO) detectable at the 5% significance level, with 90% and with 80% power, for given numbers of subjects*

	Difference in TLCO (mmol min^{-1} kPa^{-1})			
	Men		Women	
Total number of subjects (2N)	90%	80%	90%	80%
1000	0·29	0·25	0·24	0·21
750	0·33	0·29	0·28	0·24
500	0·41	0·35	0·34	0·29
400	0·46	0·40	0·38	0·33
300	0·53	0·46	0·44	0·38
200	0·65	0·56	0·54	0·46
150	0·75	0·65	0·62	0·54
100	0·91	0·79	0·76	0·66
90	0·96	0·83	0·80	0·69
80	1·02	0·88	0·85	0·73
70	1·09	0·94	0·91	0·78
60	1·18	1·02	0·98	0·85
50	1·29	1·12	1·07	0·93
40	1·45	1·25	1·20	1·04
30	1·67	1·44	1·38	1·20
20	2·04	1·77	1·70	1·47
	SD = 1·41		SD = 1·17	

The tables will be most useful for population and occupational studies. The tables are not applicable to trials in which there are paired observations, where the paired t test is used.[6] Percentages of predicted values are sometimes used for reporting lung function results. The t test is not appropriate because percentages of predicted values do not have a constant variance around the predicted value,[7] and these tables should not be used.

We thank Mrs J Gockelen and Miss C Gray for secretarial assistance.

References

1 Freiman JA, Chalmers TC, Smith H jun, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized controlled trial. *N Engl J Med* 1978;**299**:690–4.
2 Reed JF, Slaichett W. Statistical proof in inconclusive, "negative" trials. *Arch Intern Med* 1981;**141**:1307–10.
3 Altman DG. Size of clinical trials. *Br Med J* 1983; **286**:1842–3.
4 Armitage P. *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications, 1971.
5 European Community for Coal and Steel Working Party. Standardization of lung function tests. *Eur Physiopathol Respir* 1983;**19**(suppl 5).
6 Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 1981;**2**:93–113.
7 Sobol BJ, Sobol PG. Percent predicted as the limit of normal in pulmonary function testing: a statistically valid approach. *Thorax* 1979;**34**:1–3.