Sample size considerations for n-of-1 trials

Stephen Senn

Methodology and Statistics, Luxembourg Institute of Health, 1445 Strassen, Luxembourg

Summary

N-of-1 trials are trials in which patients are treated with two or more treatments on multiple occasions. They can have many different purposes and can be analysed in different frameworks. In this note five different criterion for planning sample sizes for n-of-1 trials are identified and formulae and advice to address the associated tasks are provided. Code to accomplish the tasks has been written in GenStat®, SAS® and R® and the application of the approaches is illustrated.

Introduction

N-of-1 trials[1-4] are trials in which patients are repeatedly given two or more treatments on separate occasions with the object of using these repeated episodes as a means of drawing controlled inferences about the effects of treatment. They may serve a number of different purposes. For example, repeatedly studying patients may be a way of gaining extra information that reduces the number of patients that need to be recruited. Alternatively, interest may focus on the specific effect of treatment for a given patient and this then becomes one way of delivering such information. A further use is that repeat administration of treatments in a controlled manner is a superior and principled way of identifying various components of variation. In particular, it permits a separation of random patient-by-treatment interaction from pure within-patient error[5-7]. This brings two dividends. First, it permits one to establish the degree of personal response. Second, once the components of variation have been identified, it becomes possible to produce superior 'shrunk' estimates of the effects for individual patients, using both their own response and the responses of others.

N-of-1 trials can be regarded as a subset of cross-over trials, a form of trial that has received considerable attention from statisticians. For example, there are at least five monographs devoted to the subject [8-12]. N-of-1 trials have a longer history in the field of psychology than in medicine *per se* and there are also a number of monographs devoted to their use in that field[13-16]. In that tradition, however, the emphasis has been on individual independent inferences from subject to subject, whereas what will be considered here is the situation where a number of subjects are studied repeatedly with a view to using the combined information[17-20].

In this paper, formulae for calculating sample sizes and related quantities for sets of n-of-1 trials for various purpose are presented and illustrated. Reference is made to programs in GenStat®, SAS® and R® that perform the necessary calculations. These are available on the website of the FP7-funded IDEAL project.

Assumptions and a simple model

It will be assumed that *n* patients will be repeatedly randomised in $k \geq 2$ cycles of pairs of occasions to two treatments A and B. This particular design is *relatively* common (n-of-1 trials themselves are rare) and is useful if one wishes not only to use each patient as his or her control but also achieve tight control for any local trend effects[19]. It is assumed that continuous measures are being taken, that the disease being studied is relatively stable and that carry-over can be eliminated either by using a washout or by limiting measurement to the latter part of any treatment occasion. It is also assumed that the data for each cycle for each patient can be reduced to a difference, treatment B minus treatment A (say) for each patient and that this is an efficient summary of the information for the trial. This is the case for balanced data (no missing observations per patient) that can be represented by a standard mixed model such as proposed by Araujo, Julious and Senn[20]. Where that is so, it is not necessary to build a mixed model for the original observations but instead a simpler model for the differences may be used as follows:

$$
\begin{aligned}
&d_{i(j)} = \tau_i + \varepsilon^*_{i(j)}\,,\ i = 1\cdots n,\ j = 1\cdots k,\ n \geq 2,\ k \geq 2 \\
&\tau_i \sim N\left(\mathrm{T}, \psi^2\right) \forall i \\
&\varepsilon^*_{i(j)} \sim N\left(0, 2\sigma^2\right) \forall i, j
\end{aligned}
\tag{1}
$$

Here $\varepsilon^*_{i(j)}$ are random within-cycle within-patient disturbance terms corresponding to random differences between occasions within the cycle. These terms are assumed to be identically independently distributed across cycles and patients. (The variance of $2\sigma^2$ is assigned to them to make the model compatible to one in which the disturbance terms for the original measurements from occasion to occasion are assumed independent with variance σ^2). The term τ_i is the treatment effect for patient i and this is assumed to have a common average across all patients of T and a variance of ψ^2 : the greater this variance, the greater the variation in the effect of treatment from patient to patient.

The model is completed by assuming that all error terms are independent of each other and Normally distributed.

Variances and sample size calculation

We shall consider various formulae suggested by the simple model given by (1). In order to illustrate their use, it will be instructive to have some specific parameter values in view. We shall assume that the random treatment-by-patient interaction has a variance of $\psi^2 = 1$, that the within-patient variance is $\sigma^2 = 4$, that the clinically relevant difference is $\Delta = 1$ and that various numbers of cycles ranging from $k = 2$ to $k = 15$ might be considered for design purposes. (Two is the minimum number of cycles that must be present to estimate all the components of variation. Once a set of n of 1 trials have been run they can be used to inform treatment decisions for patients with only one, or for that matter no cycles.) It is supposed that a type I error rate of $\alpha = 0.05$ two-sided is targetted and that a power of 80% is desirable so that the type II error rate is $\beta = 0.2$. The tasks to be addressed are then:

I. To find the values of n the number of patients which for given values of k will satisfy these design parameters for both a
    1. fixed effects and
    2. random effects analysis
II. To consider variances of estimates of the predicted value for a future patient studied in k cycles for both
    1. a naïve estimate (using that patient's values only)
    2. a shrunk estimate using a weighted combination of values from the given patient and previously studied patients
III. To consider the precision with which various variance components should be estimated for the purpose of personalising treatment.

Summary measures analysis of the treatment effect

Rather surprisingly, it turns out that task I.2 is easier than I.1. We use the fact that the summary measure approach for balanced data under many circumstances (including the model presented here) gives the same result as a mixed model. (See Senn, Stevens and Chaturvedi[21] for a discussion.) Here it is supposed that the data for the trial are reduced to an average treatment effect for each patient

$$\bar{d}_{i(.)} = \sum_{j=1}^{k} d_{i(j)} \Big/ k \tag{2}$$

and then the average of these is used to test the difference between treatment A and treatment B. Where this is the case, the variance of the treatment estimate is

$$\frac{\psi^2 + 2\sigma^2/k}{n} . \tag{3}$$

This is the formula considered by Zucker et al[17] and is used by them to study the trade-off between patients and cycles[4]. The mean so constructed can be used in a one sample t-test. Furthermore, the variance at the patient level, that is to say $\psi^2 + 2\sigma^2/k$ is efficiently estimated as the variance of the *n* observed summary measures and hence has $(n-1)$ degrees of freedom. Thus, standard approaches and software for sample size for the one sample t-test may be used[22]. The user simply has to decide on the number of cycles and then calculate the standard deviation at the patient level for given assumed values of ψ and σ . For example, if $k = 3$ then we may calculate that the variance of the summary statistic is $SD = \sqrt{\psi^2 + 2\sigma^2 / k}$, which for our example yields a value of $= \sqrt{1 + (2 \times 4)/3} = \sqrt{3.67} = 1.91$. Since the estimate of the variance across all patients will be based on $n-1$ degrees of freedom, this value can then be put into standard sample size software such as, for example, nQuery®, which gives the following statement:

> "When the sample size is 31, a single group t-test with a 0.050 two-sided significance level will have 80% power to detect the difference between a null hypothesis mean of 0.000 and an alternative mean of 1.000, assuming that the standard deviation is 1.910. "

Fixed effect analysis of the average treatment effect

This turns out to be trickier. This analysis is appropriate if either by hypothesis ψ^2 is assumed to be zero, or a test is made of the average treatment effect for the patients actually studied. Either of these assumptions is a reasonable justification for the limited purpose of deciding whether or not there is evidence of a difference between treatments[20]. After all, if A and B are identical for every patient, then the difference between them cannot vary from patient to patient and hence $\psi^2 = 0$. It then follows that the random treatment-by-patient interaction plays no part in the inference and can be removed from the variance component. This can be achieved in one of two simple ways. The first is to estimate the variances of the differences independently for each patient. Each such estimate has

$(k-1)$ degrees of freedom and there being n of them we obtain an overall pooled estimate based on $n(k-1)$ degrees of freedom. Alternatively, we can fit patient as a factor in the model for the nk differences so that since there are $nk-1$ degrees of freedom in total and $(n-1)$ for *patient* we are left with $(nk-1)-(n-1)=n(k-1)$ as before.

Note that since the cycle differences themselves reflect treatment effects, fitting patient effects to them effectively removes the treatment-by-patient interaction. There is an extremely strong analogy to meta-analysis where it has not been generally appreciated that the treatment-by-trial interaction is implicitly removed from a fixed effects meta-analysis[23].

Note also, that in view of the fixed effects philosophy, the random treatment-by-patient interaction ψ^2 does not contribute to the variance estimate so that the estimate of the fixed effects variance as a function of n and k is simply:

$$\frac{2\sigma^2}{nk}\ . \tag{4}$$

However, there is a problem in attempting to apply standard sample size approaches using this variance. The treatment estimate from a given patient now has a SD of $\sqrt{2\sigma^2/k}$. However, when any standard sample size program estimates the number of patients required, it implicitly assumes that the degrees of freedom for the variance are equal to $n-1$. However, as already discussed they are $n(k-1)$. Thus the degrees of freedom would be underestimated by $nk-2n+1$ This means that the true power for any sample size chosen using such an approach will be greater than sought. For example, if the value for the SD of $\sqrt{2\times 4/3}=\sqrt{2.67}=1.63$ is used with the previous parameter settings, then nQuery® proposes a sample size of 23. Such a sample size with 22 degrees of freedom for error would, indeed, have 80% power but here the degrees of freedom are $23\times(3-1)=46$ and checking these values with the non-central t-distribution yields a power of 82.0%. In fact, a calculation using the correct degrees of freedom gives a sample size of 22 patients with a power of 80.2%.

Nor can the problem be dealt with, except approximately, by using the total number of cycles nk as the target to be determined by using $\sigma\sqrt{2}$ as the relevant standard deviation. Standard sample size software would assume that the degrees of freedom for estimating σ^2 were $nk-1$ rather than $n(k-1)=nk-n$. In consequence the degrees of freedom would be overestimated by $n-1$.

In summary, the power can be checked by using the non-central t-distribution with non-centrality parameter

$$\delta=\frac{\Delta}{\sqrt{\frac{2\sigma^2}{nk}}} \tag{5}$$

and degrees of freedom $n(k-1)$. The critical value of $t_{\alpha/2,n(k-1)}$ must be established to use in connection with this. In the accompanying programs a search is made for values of n that satisfy the power requirement. One way to proceed is to start with an approximate value from the Normal distribution using the simple formula

$$n_{approx} = 2\frac{\left(Z_{\alpha/2} + Z_{\beta}\right)^2 \sigma^2}{k\Delta^2} \tag{6}$$

and then increase the value of n until the power requirement is reached. This is the approach used for cross-over trials in *Cross-over Trials in Clinical Research[10]*.

For many applications the degrees of freedom effect is unimportant. For rare diseases, however, even a difference of one patient can be of practical importance and the accompanying programs conduct a stepwise search reducing the number of patients from the naïve target number to find the minimum number that has at least the target power.

Variances for naïve estimates

If we only choose to use the information from a given patient to come to a conclusion regarding that patient, then the appropriate estimate is simply that given by(2). However, since the estimate is being used for that patient only, patient to patient variation is irrelevant and thus ψ^2 plays no role. Hence the appropriate variance is simply (3) with $\psi^2 = 0, n = 1$ and hence the standard error is given by

$$SE_{naive} = \sqrt{\frac{2\sigma^2}{k}} \tag{7}$$

For our example, given the planning parameters, the standard error of such an estimate would be $\sqrt{2\times 2^2/3} = 1.633$. Thus in planning the number of cycles to be used in n-of-1 trials and if task II.1 above is the objective, one approach is simply to set (7) equal to some target standard error and solve for k . Of course, practical matters may limit the number of cycles that can be used.

In practice when using such an estimate, however, even if only that patient's values are used to estimate the treatment effect, there is still an issue as to how the variance σ^2 itself should be estimated. In practice the degrees of freedom per patients will be very small and hence the local estimates of σ^2 will vary considerably from patient to patient[24] and there will be some value in using a pooled estimate from other patients. If this is not done, then every new patient is simply studied as if no information existed at all. This is unlikely to be sensible.

Variances for shrunk estimates

The usual general purpose of a clinical trial is to make future recommendations for patients and such recommendations will often be made on the basis of diagnosis only and without benefit of any experience of the patient on any of the possible treatment options. In other words, some average treatment effect is estimated to inform future practice. A set of n-of-1 trials clearly allows such a use also. Suppose, therefore, that a series of such trials have been conducted using a large number of patients n. A treatment estimate will have been produced of the form

$$\hat{\Delta} = \bar{d}_{.(.)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k} d_{i(j)}}{nk} \tag{8}$$

and this will have the variance given by (3). Suppose, however, that we now consider the future prediction of an effect for a patient i' not recruited in the trial and that n is large. The major uncertainty for predicting the true effect in this patient is not given by (3) (which is the variance of the overall estimate) but by ψ^2 the variance of the difference of the true effect from patient to patient.

If we now study this future patient in k cycles we shall have two estimates of the possible effect of the treatment for him or her, one $\hat{\Delta}$ based on all previous patients with variance approximately equal to ψ^2 and one of the same form as (2) based on the data for that patient only and with variance $2\sigma^2/k$. As is well known, a weighted combination of these, with weights proportional to the reciprocal of the respective variances, will provide a superior so-called *shrunk estimate* of the effect for such a patient[17, 18, 20]. The estimator is of the form

$$\hat{\tau}_{i'} = \frac{\left(2\sigma^2/k\right)\hat{\Delta} + \psi^2 \bar{d}_{i'(.)}}{\left(2\sigma^2/k\right) + \psi^2} \tag{9}$$

and the standard error of such an estimator is

$$SE_{shrunk} = \sqrt{\frac{2\sigma^2\psi^2}{k\psi^2 + 2\sigma^2}} \ . \tag{10}$$

provided that the weights can be treated as known constants. (This issue is discussed further below.)

For a future patient who contributes no information, $k = 0$ and (10) reduces to ψ . As k increases then (10) approaches the standard error for the naïve estimator given by (7). However, (10) is always less than (7) and for small ψ and k appreciable so. For the example we have considered we have

$$SE_{shrunk} = \sqrt{\frac{2\times 2^2 \times 1^2}{3\times 1^2 + 2^2}} = 0.85$$

and so approximately half the value for the naïve estimator.

For planning purpose II.1, (10) can be set equal to some target precision in order to decide on an appropriate number of cycles for a future patient.

Precision of variance components

The formulae given for the shrunk estimates assume not only that the overall treatment effect has been estimated with adequate precision but also that the variance components, ψ^2 and σ^2 have been estimated adequately. Planning purpose III addresses this.

Failure to estimate variance components adequately has two consequences: 1) the quoted standard errors will be too small, as the formula assumes that the weights to construct the shrunk estimators are known constants 2) the weights themselves will be suboptimal[25].

The precision by which the variance components have been estimated depends on the relevant degrees of freedom, $n(k-1)$ for σ^2 and, since ψ^2 cannot be estimated directly but has to be calculated from an estimate of (3) and the estimate of σ^2 two degrees of freedom are relevant for ψ^2 , $n-1$ and $n(k-1)$

However, it is really the relative values of ψ^2 and σ^2 that matter, or more precisely the relative value of the weights, $2\sigma^2/k$ and ψ^2 . The variance of the ratio of the two weights is derived in the appendix and is given by

$$Var(\hat{\rho}) = \left(\frac{k\psi^2}{2\sigma^2}+1\right)^2 \frac{2\left(n(k-1)\right)^2(nk-3)}{(n-1)\left(n(k-1)-2\right)^2\left(n(k-1)-4\right)}, \quad . \qquad (11)$$
$$n(k-1) \geq 4$$

Illustration

The use of these formulae will now be illustrated. The values of $\psi, \sigma, \Delta, \alpha$ and β previously assumed will be used. The value of k the number of cycles will be varied and various plots will be presented. Computer programs have been prepared in GenStat®, SAS® and R® to address all the calculations covered in this note and are available on the website of the IDEAL project. The figures that are included here have been produced in the GenStat® version of the program.

First, consider tasks I.1 and 1.2. These are addressed in Figure 1, which gives the desired sample size using the variance formulae given by (3) and (4) but taking care to use the appropriate degrees of freedom for the latter.

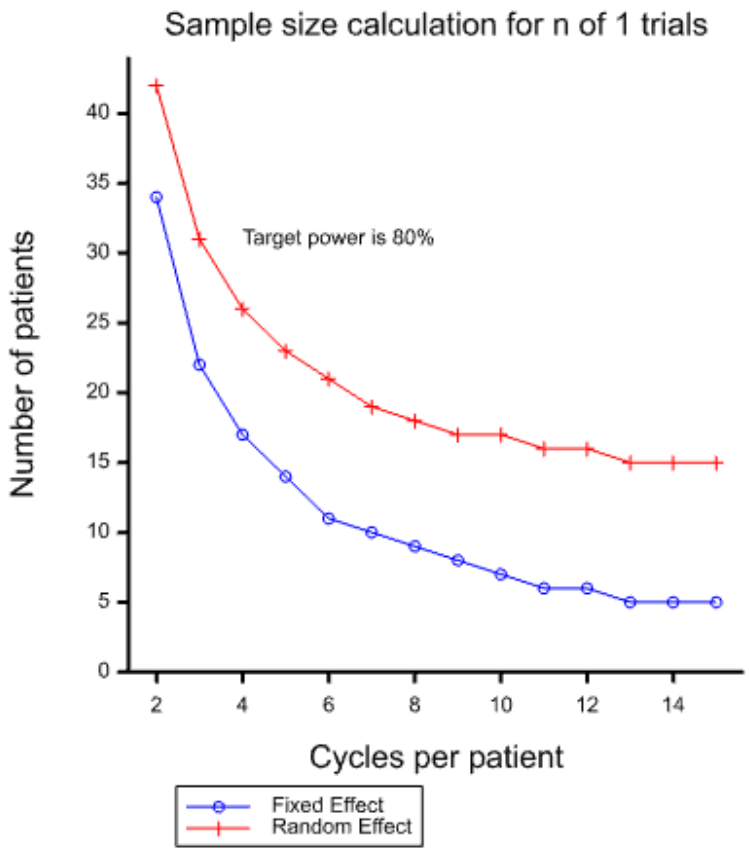

Figure 1 Sample sizes for fixed and random effects approaches

It is noticeable that as k increases the discrepancy between the sample size required for the fixed and random effect approaches increases. This is because in the latter both k and n act equally on the variance in (4) whereas in (3) only one of the components of variation is reduced by increasing k; the other is only reduced by increasing n.

As discussed in chapter 9 of *Cross-over Trials in Clinical Research[9]* an index of cost and practicality is not just the number of patients but also the total amount of patient time on treatment. One way of looking at this is in terms of the total number of cycles. This is illustrated in Figure 2, which shows that the total number of cycles varies very little with the number of cycles per patient for the fixed effects approach but continues to rise for the random effects approach.

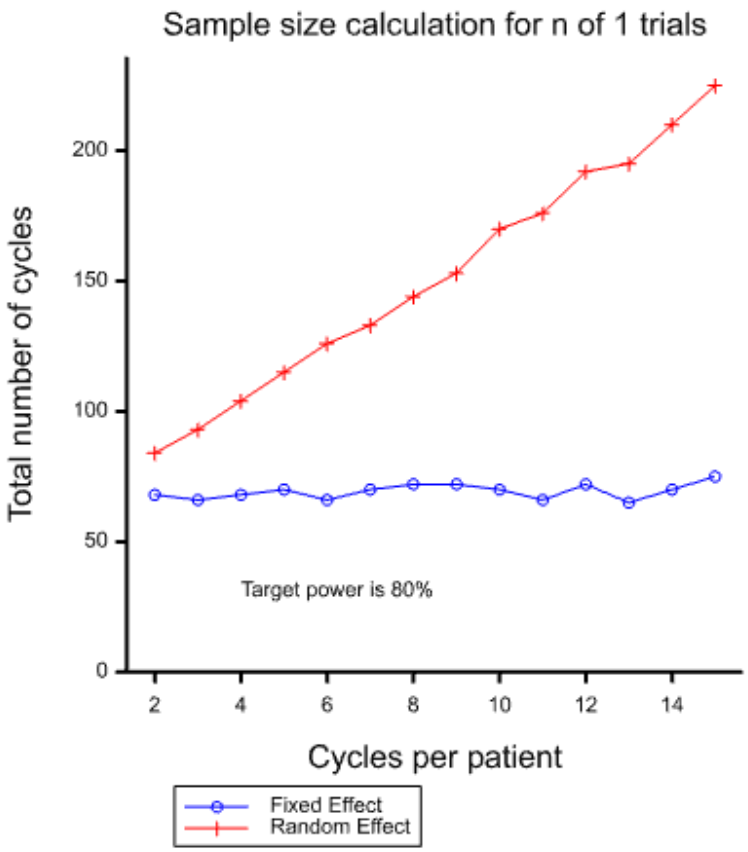

Figure 2 Numbers of cycles for two approaches to analysis

The standard errors of naïve and shrunk estimators are given in Figure 3. It can be seen that the shrunk estimate has a considerable advantage over the naïve estimate until the number of cycles is large but even then the standard error is lower. The horizontal dashed line shows the standard error of prediction making no use of the patient's own data and even this is better than the naïve estimator until the number of cycles exceeds 8.

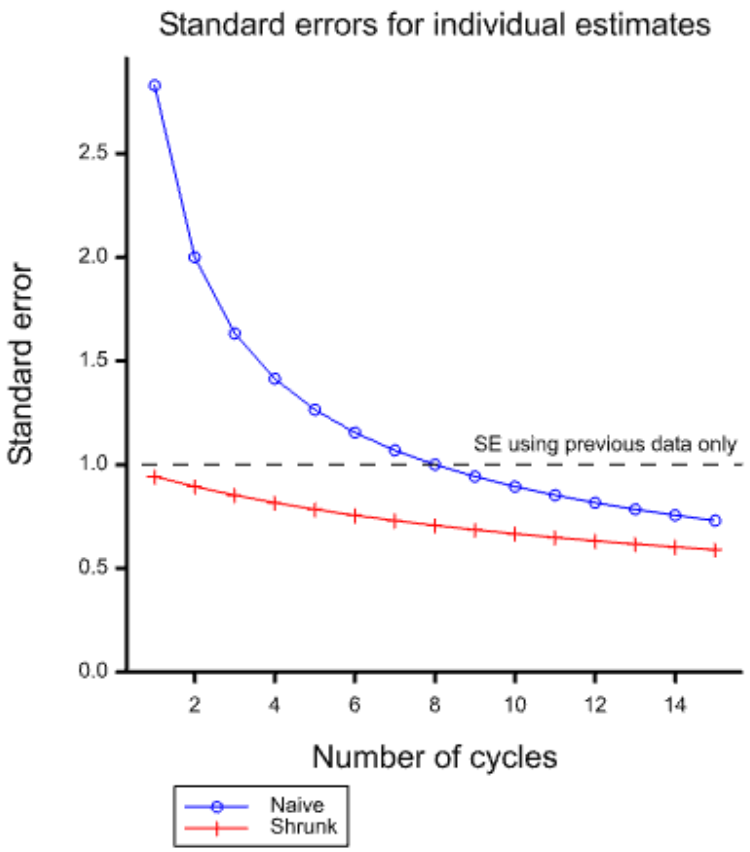

Figure 3 Standard errors of shrunk and naïve estimators for the effect for a given patient as a function of the number of cycles

Finally, Figure 4 gives the standard error of the ratio of weights for constructing shrunk estimates for the case where there are three cycles per patient. The reference line gives the ratio of the weights themselves.

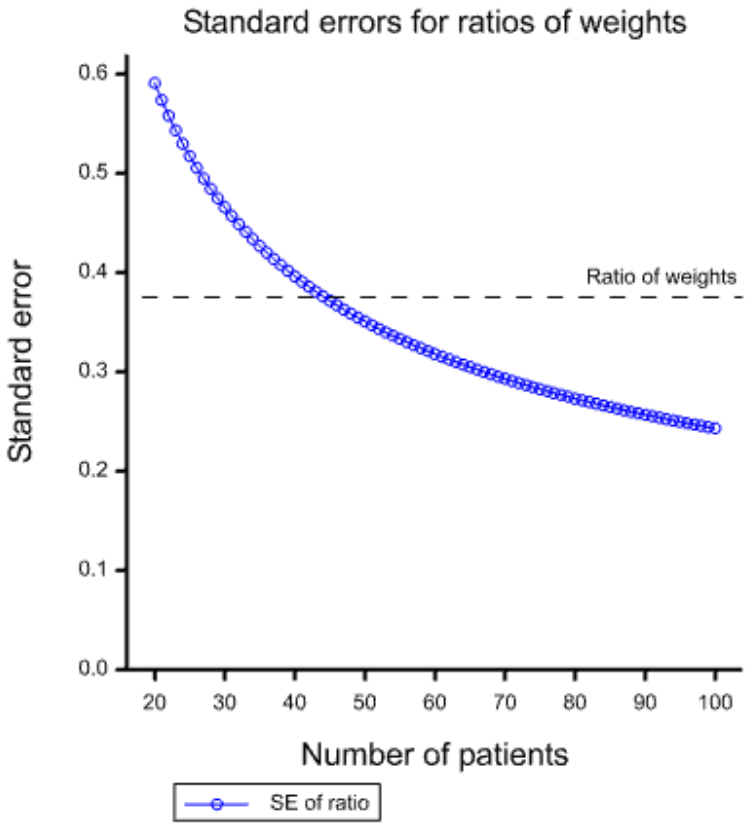

Figure 4 Standard error of the ratio of weights for constructing shrunk estimates as a function of the number of patients. The case with k = 3 cycles is illustrated.

Discussion

As within any planning problem, the key issue is to obtain some sort of impression of reasonable parameter values in advance of running the trial. Here it is best to regard Δ as a function of the disease and hence a desirable sought-for value rather than a probable one. As regards σ , again this is best regarded as a property of the disease not the treatment and previous experience may help. The most difficult of the three is ψ . As has been pointed out elsewhere[26], trialists have had a very bad track record of estimating the personal component of treatment response and in general plausible estimates for various diseases are not available simple because the sort of study necessary to estimate them, for example n-of-1 studies, have not been conducted . A pilot study may be the best approach. This raises

general design considerations beyond the scope of this paper but the reader is referred to discussion papers on this subject[27, 28].

The fomulae and the figures show that very different purposes for n-of-1 trials may require very different samples sizes. For the limited purpose of identifying treatments that can have an effect, the fixed effects approach to analysis is reasonable and this will yield the most modest requirement for sample size.

At the other end of the scale, Figure 4 paints a very gloomy picture. Large numbers of patients will be needed to estimate the components of variation used in producing shrunk estimates with adequate precision. Thus if one of the purposes of beginning a programme of n-of-1 studies is to provide improved methods of personalising treatment for individual future patients rather many patients will need to be studied in the first place. At the moment, this does not seem to be a realistic prospect. Much is talked about personalising medicine but little is done that treats components of variance seriously.

Figure 2 shows that for the fixed effects approach the total number of cycles required to be observed is fairly stable as k the number of cycles per patient increases. This may be understood as follows. Let the total number of cycles be $m = nk$. We then have that for any fixed value m (4) does not change with k. What changes are the degrees of freedom for estimating the variance which we may write as

$$n(k-1) = m\frac{k-1}{k}\ . \tag{12}$$

Thus we get some increase of degrees of freedom with increasing numbers of cycles per patient but the effect becomes relatively less important in particular because the variance of the t-distribution itself approaches 1 as the degrees of freedom increase.

Acknowledgement

This research has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant Agreement no 602552 for the IDEAL project.

Declaration of interest

I act as a consultant to the pharmaceutical industry and also for Statistical Solutions, developers of nQuery®. I maintain a full declaration of interest here
http://www.senns.demon.co.uk/Declaration_Interest.htm

References

1. Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. Controlled clinical trials. 1990;11(2):88-100. Epub 1990/04/01. PubMed PMID: 2161315.
2. Guyatt GH, Heyting AH, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 trials for investigating new drugs. Controlled clinical trials. 1990;11:88-100.
3. Keller JL, Guyatt GH, Roberts RS, Adachi JD, Rosenbloom D. An N-of-1 Service - Applying the Scientific Method in Clinical-Practice. Scandinavian Journal of Gastroenterology. 1988;23:22-9. PubMed PMID: ISI:A1988R330300004.
4. Duan N, Kravitz RL, Schmid CH. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. Journal of clinical epidemiology. 2013;66(8 Suppl):S21-8. doi: 10.1016/j.jclinepi.2013.04.006. PubMed PMID: 23849149; PubMed Central PMCID: PMCPMC3972259.
5. Dworkin RH, McDermott MP, Farrar JT, O'Connor AB, Senn S. Interpreting patient treatment response in analgesic clinical trials: implications for genotyping, phenotyping, and personalized pain treatment. Pain. 2014;155(3):457-60. doi: 10.1016/j.pain.2013.09.019. PubMed PMID: 24071599.
6. Gewandter JS, Dworkin RH, Turk DC, McDermott MP, Baron R, Gastonguay MR, et al. Research designs for proof-of-concept chronic pain clinical trials: IMMPACT recommendations. Pain. 2014. doi: 10.1016/j.pain.2014.05.025. PubMed PMID: 24865794.
7. Senn SJ. Individual Therapy: New Dawn or False Dawn. Drug Information Journal. 2001;35(4):1479-94.
8. Jones B, Kenward MG. Design and Analysis of Cross-over Trials. 3rd ed. Boca Raton: CRC Press; 2014. 412 p.
9. Ratkowsky DA, Evans MA, Alldredge JR. Cross-over experiments : design, analysis, and application. New York: Marcel Dekker; 1993.
10. Senn SJ. Cross-over Trials in Clinical Research. Second ed. Chichester: Wiley; 2002.
11. Lui KJ. Crossover Designs: Testing, Estimation and Sample Size. Chichester: Wiley; 2016.
12. Cotton JD. Analyzing Within-Subject Experiments. Mahwah, New Jersey: Lawrence Erlbaum Associates; 1998.
13. Auerbach C, Zeitlin W. SSD for R: An R package for analyzing single-subject data: Oxford University Press (UK); 2014.
14. Dugard P, File P, Todman J, Todman JB. Single-case and small-n experimental designs: A practical guide to randomization tests: Routledge; 2012.
15. Hersen M, Barlow DH. Single-case experimental designs: strategies for studying behavior change: Pergamon Press; 1976.
16. Krishef CH. Fundamental Approaches to Single Subject Design and Analysis: Krieger Publishing Company; 1991.
17. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. Journal of clinical epidemiology. 2010;63(12):1312-23. Epub 2010/09/25. doi: S0895-4356(10)00194-0 [pii]

10.1016/j.jclinepi.2010.04.020. PubMed PMID: 20863658; PubMed Central PMCID: PMC2963698.
18. Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. Journal of clinical epidemiology. 1997;50(4):401-10. Epub 1997/04/01. doi: S0895-4356(96)00429-5 [pii]. PubMed PMID: 9179098.

19. Chen X, Chen P. A comparison of four methods for the analysis of N-of-1 trials. PloS one. 2014;9(2):e87752. doi: 10.1371/journal.pone.0087752. PubMed PMID: 24503561; PubMed Central PMCID: PMC3913644.
20. Araujo A, Julious S, Senn S. Understanding variation in sets of n-of-1 trials. PloS one. 2016.
21. Senn SJ, Stevens L, Chaturvedi N. Repeated measures in clinical trials: simple strategies for analysis using summary measures. Statistics in Medicine. 2000;19(6):861-77. PubMed PMID: 0010734289.
22. Julious SA. Tutorial in biostatistics - Sample sizes for clinical trials with normal data. Statistics in Medicine. 2004;23(12):1921-86. PubMed PMID: ISI:000221805600007.
23. Senn SJ. The many modes of meta. Drug Information Journal. 2000;34:535-49.
24. Senn SJ. Letter to the Editor: in defence of the linear model. Controlled clinical trials. 2000;21:589 - 92.
25. Senn S. Various varying variances: The challenge of nuisance parameters to the practising biostatistician. Stat Methods Med Res. 2015;24(4):403-19. doi: 10.1177/0962280214520728. PubMed PMID: 24492794.
26. Senn SJ. Mastering variation: variance components and personalised medicine. Stat Med. 2016;35(7):966-77. doi: 10.1002/sim.6739. PubMed PMID: 26415869.
27. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. Journal of evaluation in clinical practice. 2004;10(2):307-12.
28. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. BMC medical research methodology. 2010;10(1):1.
29. Evans M, Hastings N, Peacock B. Statistical Distributions. Chichester and New York: John Wiley and Sons; 1993.

Appendix

From (9) we have that the ratio ρ of the true weights for constructing the shrunk estimator would be

$$\rho = \frac{\psi^2}{2\sigma^2/k} = \frac{k}{2}\frac{\psi^2}{\sigma^2} \ . \tag{13}$$

Now let the estimate of the variance within σ^2, be V_W and of the total variance $\psi^2 + 2\sigma^2/k$ be V_T . Substitute V_W for σ^2 and $V_T - 2V_W/k$ for ψ^2 in (13) to obtain an estimate of the ratio of weights $\hat{\rho}$ of

$$\hat{\rho} = \frac{k}{2}\frac{V_T}{V_W} - 1 \ . \tag{14}$$

For the purpose of estimating the variance, the constant term -1 in (14) may be ignored so from now on we may work with $\hat{\rho}^* = \hat{\rho} + 1$.

Now we write

$$\hat{\rho}^* = \left\{ \frac{V_T \Big/ \left(\psi^2 + \frac{2\sigma^2}{k} \right)}{V_W/\sigma^2} \right\} \left(\frac{k\psi^2}{2\sigma^2} + 1 \right) . \tag{15}$$

This may seem unnecessarily complicated but the point is that the term in the curly parentheses is an $F_{n-1,n(k-1)}$ statistic.

In general, the variance of a random variable distributed $F_{\nu,\omega}$ is

$$Var\left(F_{\nu,\omega}\right) = \frac{2\omega^2\left(\nu+\omega-2\right)}{\nu\left(\omega-2\right)^2\left(\omega-4\right)}, \quad \omega > 4 \ . \tag{16}$$

(See, for example, Evans, Hastings and Peacock[29] p69.) Hence, by substituting $\nu = n-1, \omega = n\left(k-1\right)$ in (16) and multiplying by the square of rightmost term of(15) we obtain the variance of $\hat{\rho}^*$ and hence of $\hat{\rho}$ as

$$Var\left(\hat{\rho}\right) = \left(\frac{k\psi^2}{2\sigma^2} + 1 \right)^2 \frac{2\left(n\left(k-1\right)\right)^2\left(nk-3\right)}{\left(n-1\right)\left(n\left(k-1\right)-2\right)^2\left(n\left(k-1\right)-4\right)} \ . \tag{17}$$

Note that for the variance given by (17) to be defined we must have $n > 1$ and $n > \frac{4}{k-1}$ and the minimum value for the number of cycles for estimation of components of variation to be possible is

$k = 2$ which would yield a minimum value of $n = 5$. Since n must be integer, for $k = 3$ we have $n = 3$ as a minimum and for $k = 4$, we have $n = 2$. The product nk for these three cases is 10, 9 and 8 total cycles respectively.