PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2005; **4**: 287–291
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pst.185

Sample size of 12 per group rule of thumb for a pilot study

CONSULTANTS FORUM

Steven A. Julious*,†
Medical Statistics Unit, Institute of General Practice and Primary Care, University of Sheffield, Community Sciences Centre, Northern General Hospital, Sheffield, UK

When designing a clinical trial an appropriate justification for the sample size should be provided in the protocol. However, there are a number of settings when undertaking a pilot trial when there is no prior information to base a sample size on. For such pilot studies the recommendation is a sample size of 12 per group. The justifications for this sample size are based on rationale about feasibility; precision about the mean and variance; and regulatory considerations. The context of the justifications are that future studies will use the information from the pilot in their design.

Keywords: *sample size; pilot study*

1. INTRODUCTION

When designing a clinical trial an appropriate justification for the sample size should be provided in the protocol. This justification could be based on formal power calculations [1] or on other considerations such as the precision of the estimates of interest [2]. However, there are a number of settings when designing a pilot investigation where there is no prior information upon which to base the sample size. For example, in phase I the study could be a bioavailability study for a new chemical entity, while for a later phase the study could be with a novel endpoint or in a previously unstudied group of patients (for the compound). In these situations the intention is that later more definitive studies may be carried out and the recommendation would be a sample size of 12 per group as being appropriate.

The fact that future investigations are (or may be) planned is important, for no study is an island [3], and the justifications described are in this context. Three reasons for justifying a sample size of 12 per group will be given based on: feasibility; gains in the precision about the mean and variance; and regulatory considerations.

The assumption through the paper is that the endpoint of interest is anticipated to take a Normal form.

2. RATIONALE

2.1. Reason 1: Feasibility

In the design of a parallel group trial a sample size of 12 per group is a good round number. It is divisible by 2, 3, 4 and 6 and so facilitates the

*Correspondence to: Steven A. Julious, Medical Statistics Unit, Institute of General Practice and Primary Care, University of Sheffield, Community Sciences Centre, Northern General Hospital, Herries Road, Sheffield S5 7AU, UK.
†E-mail: s.a.julious@sheffield.ac.uk

Pharmaceutical STATISTICS

setting of a variety of block sizes. You could have block sizes of 2, 3, 4, 6, 8 or 12 (or even 24) if you had two groups and a total sample size of 24.

For multi-period cross-over investigations, common with early phase I investigations, with a total sample size of 12 you could have 2, 3, 4 and 6 (or even 12) period cross-over trials with balanced Williams squares designs [4]. For a 5 period cross-over trial you need a sample size that is a divisible by 10, and so feasibility here may reduce the sample size to 10.

2.2. Reason 2: Precision about the mean and variance

2.2.1. Precision about the mean

Obviously the greater the sample size the smaller the standard error and as a consequence the greater the precision about the mean difference – as assessed by its confidence interval. A two-sided confidence interval for a parallel group trial is defined as

$$\bar{x}_A - \bar{x}_B \pm t_{1-\alpha/2,2n-2}\sqrt{2}s/\sqrt{n} \quad (1)$$

where $\bar{x}_A$ and $\bar{x}_B$ are the means on treatments A and B, s^2 is the common variance for the two groups and n is the sample size in each group. The situation we are considering here is to assess, with a finite sample size, what gain in precision we have for every unit increase in the sample size per group. This could be assessed using the right-hand side of (1) and the expression

$$\frac{t_{1-\alpha/2,2n-2}\sqrt{2}s}{\sqrt{n}} - \frac{t_{1-\alpha/2,2(n+1)-2}\sqrt{2}s}{\sqrt{n+1}} \quad (2)$$

From (2) Figure 1 can be derived. This was estimated assuming a unit variance ($s^2 = 1$) and that a two-sided 95% confidence interval would be used in the planned trial.

The point associated with '4' on the x-axis gives the increase in precision of a sample size of over one of 3. The point associated with 5 gives the gain over 4, and so on. Therefore, from Figure 1 it seems that for small sample sizes there is a marked gain for each increase of 1 in the sample size per group, but that the gains are less pronounced by the point where the sample size has reached 12. van Belle undertook similar calculations and also recommended a sample size of 12 per group for a pilot investigation [5].

Note that the y-axis in Figure 1 (as well as in Figures 2 and 3), although calculated using a unit variance, could be considered as a multiple of the estimated standard deviation.

Equivalent to (2), the following can be used for cross-over trials to estimate the gain in precision for each increase in the total sample size of 1:

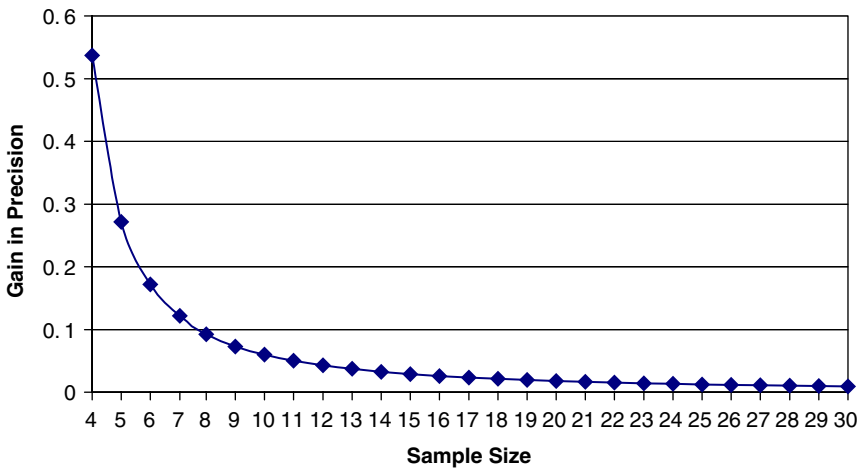

Figure 1. Gain in precision for each increase of 1 in the sample size per arm for a parallel group trial.

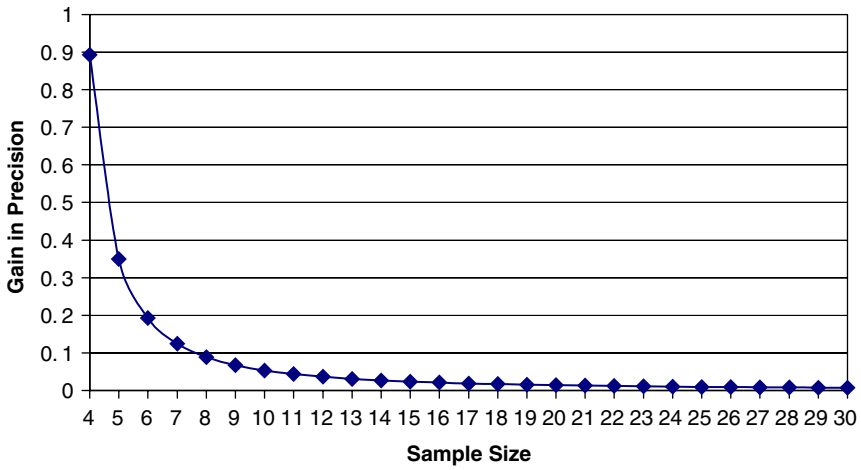

Figure 2. Gain in precision for each increase of 1 in the total sample size for a cross-over trial.

$$\frac{t_{1-\alpha/2,n-2}\sqrt{2}s_{\mathrm{w}}}{\sqrt{n}} - \frac{t_{1-\alpha/2,(n+1)-2}\sqrt{2}s_{\mathrm{w}}}{\sqrt{n+1}} \quad (3)$$

where s_{w}^2 is the within-subject estimate of the variance. The degrees of freedom associated with the *t*-statistic here are given assuming the trial is a two-period cross-over study analysed using an analysis of variance with period fitted in the model. From (3) Figure 2 can be derived. Similar to parallel group trials, after a sample size of 12 the gains in precision become less pronounced.

2.2.2. *Precision about the variance*

As well as quantifying an estimate of possible effect, a pilot study is important to provide an estimate of the variance as this variance may be used in a formal sample size calculation in a subsequent study. One way of assessing the precision of the variance would be to determine the sensitivity of the future formally powered study to the estimate of the (now assumed unknown) variance. A high plausible value of the variance could be used to assess the sensitivity of a study by utilizing the degrees of freedom for the variance estimate and the chi-squared distribution – a detailed description of the methodologies is given by Julious [6]. This high plausible value could be estimated using the upper one-tailed 95th percentile for the variance from the formula

$$s^2(95) < \frac{df}{\chi^2_{0.95,df}}\, s^2 \quad (4)$$

where s^2 is estimated in the pilot study and *df* is the degrees of freedom for s^2 from this study [6]. The value of *df* is directly related to the sample size in the study.

What is therefore required from a pilot study is a sample size sufficiently large to have appropriate degrees of freedom for s^2 in a sensitivity analysis in the future study.

The process in the formal sample size calculation in the future study would be as follows:

1. Obtain an estimate of the variance (s^2) from your pilot study.
2. Calculate the sample size from this variance.
3. With this sample size determine what would be the sensitivity of the study to the assumptions about the variance. This is assessed as a loss in power, to a high plausible value for s^2 taken from (4).

Similar to (2) and (3), from (4) the following could be used to assess the gain in precision

Pharmaceutical STATISTICS

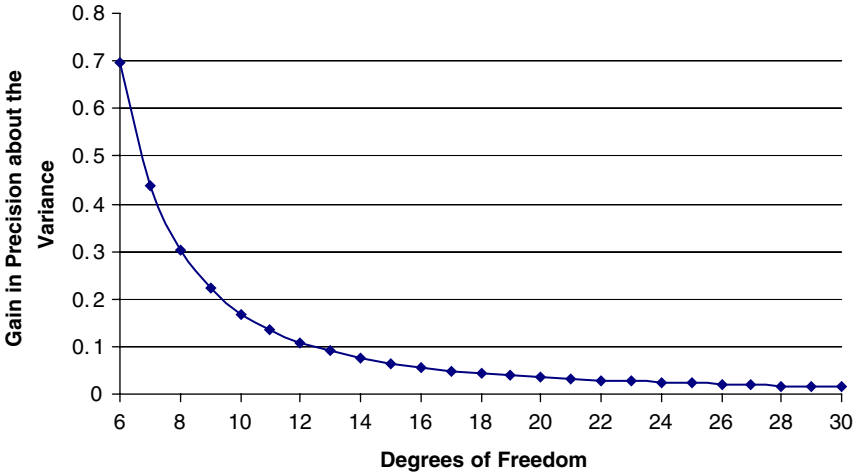

Figure 3. Gain in precision for each increase of 1 in the degrees of freedom.

around the variance for each gain in the degrees of freedom,

$$\frac{df}{\chi^2_{0.95,df}} - \frac{df+1}{\chi^2_{0.95,df+1}} \quad (5)$$

and from (5) Figure 3 can be constructed. In Figure 3 the asymptote seems to appear when the degrees of freedom exceeds around 20.

Another way to utilize (4) is to consider, if a study was designed with 90% power, what degrees of freedom would be required to ensure that you would have at least 50% power, even with a high plausible value for the variance (as assessed through the 95th percentile) from a sensitivity analysis. With this reasoning Table I was constructed. Note that 50% power is taken due to the small sample sizes of the pilot investigation.

Table I can only be taken as a rough guide and not a generic recommendation. For the superiority, non-inferiority and equivalence studies the assumption is that the future study will be a parallel group study, whilst for the bioequivalence the assumption is that the design will be a cross-over. For the non-inferiority, equivalence and bioequivalence studies the assumption is that a future study will be designed assuming no true difference between treatments. Obviously this table will not hold if the assumptions for future studies are different from those assumed in the table.

Table I. Degrees if freedom required to ensure the 95th percentile for the variance has 50% power.

Type of trial	Degrees of freedom
Superiority	9
Non-inferiority	9
Equivalence	20
Bioequivalence	16

Table I is quite informative. For equivalence and bioequivalence studies it seems that greater degrees of freedom are required. However, for a bioequivalence study you could imagine the variance to be estimated from a multi-period trial – for example, a three-period, three-treatment study with 12 subjects in total would be anticipated to have 20 degrees of freedom for the residual error. For an equivalence trial you could similarly imagine the variance to be estimated from a parallel group trial (with 12 per arm). Hence, the requisite degrees of freedom for Table I you could suppose to be attainable.

Consistent with Table I other work by Birkett and Day on internal pilots is of interest [7]. These authors recommend a minimum of 20 degrees of freedom for the variance. The situation of internal pilots could be considered to be analogous to the situation of a pilot study internal to a wider

development plan and so this work could be borne in mind.

2.3. Reason 3: Regulatory considerations

For general guidance for a sample size calculation ICH simply states [8]: 'The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed. This number is usually determined by the primary objective of the trial. If the sample size is determined on some other basis, then this should be made clear and justified.' Hence if a study is a pilot with the sample size based on feasibility it should be expressly stated as such in the protocol. If based on feasibility you may wish to calculate the precision for the confidence interval(s) around the primary endpoint(s) and include this as part of the justification for the sample size [1, 2].

For pilot bioavailability and food investigations there is some regulatory justification for choosing a sample size of 12 – justification that could be extended to other types of phase I studies. For food effect studies FDA Guidance states [9]: 'A minimum of 12 subjects should complete the food-effect BA and fed BE studies', hence, that you should over-recruit to ensure you have 12 subjects for the assessment of food effect studies. Similarly, for assessing bioavailability and bio-equivalence FDA Guidance states [10]: 'A minimum number of 12 evaluable subjects should be included in any BE study'. However, the most interesting justification comes in general considerations guidelines on bioavailability/bioequivalance from the FDA [11]: 'If the sponsor chooses, a pilot study in a small number of subjects can be carried out before proceeding with a full BE study. The study can be used to validate analytical methodology, assess variability, optimize sample collection time intervals, and provide other information. ... A pilot study that documents BE can be appropriate, provided its design and execution are suitable and a sufficient number of subjects (e.g., 12) have completed the study.'

3. DISCUSSION

In this paper rationales have been given as to an appropriate sample size for a pilot investigation. The justification was centred on both feasibility and the precision around the estimates to be used to design future studies. It is recommended that a minimum 12 subjects per group be considered for pilot studies.

REFERENCES

1. Julious SA. Tutorial in Biostatistics: Sample sizes for clinical trials with Normal data. *Statistics in Medicine* 2004; **23**:1921–1986.
2. Julious SA, Paterson SD. Sample sizes for estimation in clinical research. *Pharmaceutical Statistics* 2004; **3**:213–215.
3. Julious SA, Swank D. Moving statistics beyond the individual clinical trial: applying decision science to optimize a clinical development plan. *Pharmaceutical Statistics* 2005; **4**:37–46.
4. Jones B, Kenward MG. *Design and analysis of cross-over trials* (2nd edn). Chapman & Hall/CRC: Boca Raton, FL, 2003.
5. van Belle G. *Statistical rules of thumb*. Wiley: Chichester, 2002.
6. Julious SA. Designing clinical trials with uncertain estimates of variability. *Pharmaceutical Statistics* 2004; **3**:261–268.
7. Birkett MA, Day SJ. Internal pilot studies for estimating sample size. *Statistics in Medicine* 1994; **13**:2455–2463.
8. ICH. Guidance for industry E9. Statistical principle for clinical trials. September 1998. http://www.fda.gov/cder/guidance/ICH_E9-fnl.PDF (accessed 2 December 2004).
9. FDA. Guidance for industry. Food-effect bioavailability and fed bioequivalence studies. December 2002. http://www.fda.gov/cder/guidance/5194fnl.pdf (accessed 25 November 2004).
10. FDA. Guidance for industry. Statistical approaches to establishing bioequivalence. January 2001. http://www.fda.gov/cder/guidance/3616fnl.pdf (accessed 25 November 2004).
11. FDA. Guidance for industry. Bioavailability and bioequivalence studies for orally administered drug products – general considerations. March 2003. http://www.fda.gov/cder/guidance/5356fnl.pdf (accessed 25 November 2004).

