UNIVERSITY OF LINCOLN

CLINICAL GOVERNANCE - STANDARD OPERATING PROCEDURE
SAMPLE SIZE CALCULATIONS
CG-QMS SOP CGS3

Version Final 1.0 Date 01 May 2021

Effective Date: 01 August 2021

Next review: 2 years

Author:	Professor Graham Law Professor of Medical Statistics
Approved by: Signature:	UREC See original

Version History	**Reason for change**

NOTE: All SOPs are subject to regular review.

Please ensure that the version of this SOP is the most up-to-date.

OUT OF DATE DOCUMENTS MUST NOT BE USED AND HARD COPIES SHOULD BE DESTROYED

1 PURPOSE

This SOP should be used as the outline of statistical principles for calculating the sample size for a study, with consideration of the design and analysis.

2 SCOPE

This Standard Operating Procedure will be followed by statisticians/analysts working on research projects coordinated by the Lincoln Clinical Trials Unit.

3 BACKGROUND

The number of participants in a clinical study, known as the 'sample size', should be large enough to provide a reliable answer to the hypothesis being tested.

The sample size should be calculated usually to examine the primary objective. Drop-out will be needed to estimate the required sample size.

Some but not all points are relevant to observational studies.

Sample size calculations should be considered at an early stage of the planning for a research project.

The required sample size for specific study designs and statistical analyses mentioned below can be computed using software such as Stata, G Power calculator, and others. The line of code and/or the specific parameters used for computation such as α, power, and estimated effect size should be specified.

Members of the statistical team should be listed on the RF CG08-RF01 Study Delegation Log.

4 CROSS REFERENCES

4.1 CG-QMS SOP CGS2Statistical Analysis Plans

4.2 CG-QMS SOP CGD6 Randomisation

4.3 CG-QMS RF CG08-RF01Study Delegation Log

5 PROCEDURE

5.1 The statistician (or someone with relevant knowledge, skills and experience) must define:

- Study design being used.
- Primary efficacy outcome.
- The null hypothesis.
- The alternative hypothesis, typically, a minimal clinically relevant effect will need to be agreed upon. There should be justification that this is achievable.
- α (alpha) level: probability of Type I error – incorrectly rejecting the null hypothesis. Preferably two-sided and justified when not. Typically 5%, where different, justification required.
- β (beta) level: probability of Type II error - failing to reject the null hypothesis. 10% or 20% typically.
- Depending on the study design, outcome and exposure/intervention being used, the following may be required: mean, variances, intra-correlation coefficients, response rates, difference to detect, compliance and completion.
- Adjustment will be made for the estimated dropout rate:
  - (inflating by 1/(1-dropout rate))
- Adjustment will be made for estimated non-compliance:
  - (inflating by 1/(1-total non-compliance rate))

This information will allow the calculation of the sample size.

Avoid post-hoc calculation of sample size.

OBSERVATIONAL: CROSS-SECTIONAL, PREVALENCE

5.2 When calculating the sample size required for a cross-sectional survey the sample size is estimates as:

$$Sample\ size = \frac{{z_{1-\propto/2}}^2 SD^2}{d^2}$$

Where ${z_{1-\propto/2}}^2$ is the standard normal variate ie. 1.96,

SD is the standard deviation of the variable,

d is the absolute error or precision.

This same thing can be estimated for proportion in the population. For example:

$$Sample\ size = \frac{{z_{1-\propto/2}}^2 p(1-p)}{d^2}$$

Where ${z_{1-\propto/2}}^2$ is the standard normal variate ie. 1.96

P is the estimated proportion,

d is the absolute error or precision.

See for further details: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3775042/

EXPERIMENTAL: FEASIBILITY STUDY

5.3 NIHR guidance (http://www.rds-sw.nihr.ac.uk/dloads/rfpb_feasibility_trials_guidance.pdf) referenced Julious 2005[1] and said the usual sort of power calculation: the sample size should be adequate to estimate the critical parameters (e.g. recruitment rate) to the necessary degree of precision.

5.4 Feasibility study prior to a two-parallel-balanced-group superiority RCT. For normally distributed outcomes, the relative gain in precision of the pooled standard deviation. Teare and colleagues found that the optimal size for pilot studies used to inform sample size calculations was between 70 and 120 participants, depending on the nature of the outcome data. Feasibility studies and, to some extent, pilot studies may contain multiple sub-studies of varying sizes. For example, a particular feasibility study might involve: a survey of up to 40 clinical staff to assess their theoretical willingness to recruit patients and to follow trial procedures; interviews with up to 20 patients to assess their theoretical willingness to participate in a future trial and their views on options for intervention delivery and outcome measurement; and a review of clinical data, with no participants, to assess data accuracy and completeness.

EXPERIMENTAL: RANDOMISED CONTROLLED TRIAL

5.5 Sample size will need to be defined as either: (i) full analysis set (all randomised participants), or (ii) intention to treat or modified intention to treat, or (iii) per protocol set (participants that may be evaluated).

TWO PARALLEL GROUPS, CONTINUOUS OUTCOME

5.6 Assuming the outcome is approximately Normally distributed, and a simple 2 sample t-test will be used to compare the treatment groups, the sample size calculation will follow standard methods that require specification of:

- The smallest difference in means of clinical importance (e.g. difference in average blood pressure)
- Standard deviation(s) (SD) of the measures

[1] Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics 2005, 4:287–291.

- For non-Normally distributed data where a Mann-Whitney test is planned, a conservative estimate of sample size may be obtained by dividing the t-test sample size by 0.864 (Randle & Wolfe, 1979)[2].

Continuous outcome with baseline measurements requiring ANCOVA an adjustment should be made to the sample size calculation. This requires an estimate of the correlation between repeated measurements of the primary outcome.

TWO PARALLEL GROUPS, BINARY OUTCOME

5.7 For a binary outcome where a simple Chi squared test is planned for analysis the following information will be required to calculate sample size

- smallest effect size of clinical importance (e.g. difference in proportions)
- estimate of the prevalence of the disease in the control group

TWO PARALLEL GROUPS, SURVIVAL OUTCOME

5.8 For a survival outcome where the basic analysis will involve a log rank test with an assumption of proportional hazards, the following information is required for the sample size calculation:

Either

- anticipated values for the proportions surviving in the two groups at a chosen time point (e.g. 1 year)

Or

- anticipated values for the proportion surviving in the comparison group and an estimate of the hazard ratio.

Or

- anticipated values for the proportion surviving in the comparison group and estimates of the median survival times in the two groups (assuming an exponential distribution)

MORE THAN 2 PARALLEL GROUPS

5.9 In the case of more than two treatment groups, the sample size for a comparison of two groups should be used and multiplied up to allow for the number of parallel groups planned. An adjusted P-value should be used in this calculation (e.g. based on Bonferroni) to allow for multiple testing.

EXPERIMENTAL: RANDOMISED CROSS-OVER TRIAL

CONTINUOUS OUTCOME

5.10 Assuming the outcome is approximately Normally distributed the sample size calculation can be carried out based on a paired t-test. This requires specification of:

- the smallest clinically important difference in the outcome
- the SD of the differences between repeated measurements for the outcome

BINARY OUTCOME

5.11 Assuming a McNemar's test is appropriate for the primary analysis, the sample size calculation will require estimates of:

- proportion of discordant pairs
- odds ratio (of responding positively on treatment B and negatively on treatment A compared with positively on treatment A and negatively on treatment B)

EXPERIMENTAL: CLUSTER RANDOMISED TRIALS OR TRIALS WITH REPEATED MEASUREMENTS

[2] Randle R., Wolfe D. Introduction to the theory of nonparametric statistics, Wiley, New York 1979.

5.12 The sample size obtained for a parallel group study with a single outcome measurement should be inflated by the design effect. This requires estimates of the average cluster size and the intra cluster correlation coefficient.

OTHER DESIGNS AND CONSIDERATIONS

INTERIM ANALYSES

5.13 In trials where one or more interim analysis is planned, an appropriate adjustment to the significance level should be made in the sample size calculation. In addition, details should be given of planned stopping rules.

MULTIPLE PRIMARY ENDPOINTS

5.14 Where multiple primary endpoints are planned the sample size should be calculated for each using a significance level that reflects the number of comparisons to be made (e.g. using the Bonferroni method). The largest sample size should then be planned for use in the study.

6 FLOW CHART

None required.