CLINICAL INVESTIGATION

EDITORIAL

Clin. Invest. (2012) 2(7), 655–657

Small sample sizes in clinical trials: a statistician's perspective

Lucinda Billingham*[1,2], Kinga Malottki[1] & Neil Steven[2,3]

Small sample sizes can occur in Phase III clinical trials, either by design because the disease is rare or as a result of early closure due to recruitment failure. In either case there is a need to think differently about the statistical analysis, as the more traditional approaches may be problematic. In the case of a rare disease, there is an opportunity to plan the statistical analysis to account for the expected small numbers of patients; whilst in the failed trial, there may be a need to change the statistical analysis plan in order to maximize the usefulness of the information provided by the unexpected smaller number of patients. Clinicians have to make difficult treatment decisions for their patients on a daily basis and although small sample sizes are not ideal, there are ethical arguments to consider. Patients with rare diseases have the right for treatment decisions to be based on some level of unbiased evidence and in a failed trial it is ethical to analyze the data in such a way that the data can still aid decisions and, thereby, provide some return for the investment made by patients and funders.

Traditionally, Phase III trial designs are based on hypothesis testing. Typically, this approach tests the null hypothesis of no treatment effect against the alternative hypothesis that there is a treatment effect. The size of the trial is based on maximizing the chances of making a correct conclusion from the trial data; in particular, trials are designed to have a good chance (usually 90%) of rejecting the null hypothesis (at a 5% significance level) when a prespecified minimum clinically relevant treatment effect truly exists, a feature known as power. The problem with this approach in a trial with small sample size is that the analysis will be underpowered and the trial is unlikely to make the correct conclusion. Less conventional methodological approaches are supported if they help to improve the interpretability of trial results [1,2].

Clinical trials aim to gather unbiased evidence regarding a treatment effect but, rather than trying to provide a definitive answer through hypothesis testing, an alternative view is to consider trials as a way of reducing uncertainty about the size of a treatment effect. If one starts from the premise that there is considerable uncertainty regarding this unknown quantity, then data from even small numbers of patients in a well-designed clinical trial will make steps towards reducing that uncertainty. This improved information will help clinicians in the treatment decisions that they need to make with their patients. This alternative statistical view lends itself to using a Bayesian approach to analysis [3,4]. This was the view and methodology proposed by one of the earliest papers to discuss designing trials in rare diseases [5] and the Bayesian approach was also advocated at that time more generally in relation to small clinical trials [6]. We support this Bayesian approach, but there are issues in its implementation that we would like to highlight in this editorial.

Keywords: Bayesian analysis • Merkel cell carcinoma • rare diseases • recruitment failure • standardized likelihood

[1]MRC Midland Hub for Trials Methodology Research, University of Birmingham, Birmingham, UK
[2]Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK
[3]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
*Author for correspondence:
E-mail: l.j.billingham@bham.ac.uk

ISSN 2041-6792 10.4155/CLI.12.62

The Bayesian approach allows external and subjective information about the size of the treatment effect, expressed as a prior probability distribution, to be combined with trial evidence to give a posterior probability distribution for the size of the treatment effect. This approach ensures that the trial is reducing the uncertainty about the treatment effect from a level that already exists. For example, if there is relatively strong prior evidence that the treatment is effective, either through subjective belief or by summarizing existing evidence, then a small trial that supports this may be all that is required to change clinical practice. However, the situation becomes more complex if the trial data and prior evidence conflict. An additional key advantage with the Bayesian approach is that the results from a trial can be expressed in terms of direct probabilities of the treatment effect being a certain size. For example, the sort of result that one would be able to conclude in terms of a survival outcome is that, given prior evidence and the trial data, there is a 70% chance that the treatment truly reduces the hazard of death by at least 10% (i.e., hazard ratio <0.9). In small studies, this type of reporting could be used practically by clinicians in discussion with patients and enable evidence-based treatment decisions, whilst a non-significant result from hypothesis testing would simply be regarded as inconclusive or, at worst, evidence of no treatment effect.

"...although small sample sizes are not ideal, there are ethical arguments to consider."

Further to the proposal by Lilford and colleagues, a strategy was developed for designing trials to evaluate interventions in rare cancers, specifically in terms of survival time as an outcome measure [7–9]. It proposed a methodology for creating a prior distribution from existing evidence. The strategy suggests searching the literature for all evidence relating to a proposed trial, even including studies where there are only tentative similarities in terms of type of cancer, treatment and end points, and including all levels of evidence from randomized controlled trials to single case study reports. This evidence can then be combined into a prior distribution for the treatment effect with weights allocated in relation to pertinence, validity and precision. In principle this idea is sensible, but in practice such an approach is problematic, as discovered when applying this methodology to design a trial of adjuvant chemotherapy in stage I-III Merkel cell carcinoma. Such broad search strategies can produce large numbers of potentially relevant papers and in rare diseases it is unlikely that any of these will be high-level evidence. From around 27,000 references identified in searches related to the planned Merkel cell carcinoma trial, approximately 1000 were found to be potentially relevant and the majority were case studies with a single-arm study as the best level of evidence. Reviewing these and extracting data is extremely time-consuming and estimating hazard ratios from such studies without direct treatment comparisons is not straightforward. More importantly, such evidence is potentially so biased that the prior probability distribution would not be believable. In addition, the poor quality evidence is allocated very low weights in the strategy, 0.3 for single arm study to 0.05 for case study compared to 1 for a randomized controlled trial and, therefore, despite the large effort needed to extract and combine such information, it ends up contributing very little to the prior. One has to question the value of undertaking such a strategy.

"If one starts from the premise that there is considerable uncertainty regarding this unknown quantity, then data from even small numbers of patients in a well-designed clinical trial will make steps towards reducing that uncertainty."

Given this difficulty in producing an evidence-based prior and the fact that many clinicians find it difficult to accept the inclusion in the analysis of a prior based on subjective beliefs, we need to consider the alternatives. Actually, the Bayesian approach can still be applied by using a noninformative prior distribution. This is effectively a uniform probability distribution that reflects the fact that every size of treatment effect is equally likely, because there is no evidence to believe otherwise. An analysis with this type of prior ensures that the posterior probability distribution for the treatment effect is totally dominated by the data from the trial. Technically, the distribution coincides with the likelihood, which is a probability function that shows how strongly the data support every possible value of the treatment effect. When combined with a noninformative prior, this is often referred to as a 'standardized likelihood' [4] and such an approach could be called a 'likelihood-based Bayesian analysis'. The reason that such an approach is still useful is that, as specified earlier, it enables the results to be expressed in terms of direct probabilities of the treatment effect size being within a certain range but this time based purely on the results from the trial. This type of approach can be effective in maximizing the value of the information from a trial that has failed to recruit. In terms of rare diseases, if such an analysis is planned, then a sample size can be chosen that is feasible and ensures that the posterior probability distribution has an acceptable level of uncertainty that will enable clinical decisions. This approach using standardized likelihood has been suggested before as a useful approach to presenting trial results to clinicians [10,11], not only

for small sample sizes but as a companion analysis to traditional approaches in trials of any size. Although the concept is appealing, noninformativeness is not necessarily straightforward [11].

Finally, there is an inclination to use a simple approach, called conjugate analysis, to estimate the posterior distribution, as proposed in the strategy by Tan and colleagues [7]. Essentially, this means that the posterior probability distribution is estimated simply by combining values of the parameters that define the distributions of the prior and likelihood, weighted according to the amount of information in each. Unfortunately, the exact scenario where this simplistic approach to Bayesian analysis may fail is when there is a small sample size. In the long run, randomization in a clinical trial will produce balanced patient groups, but with small sample sizes there is a high chance of imbalance in terms of potential prognostic factors and, therefore, these need to be adjusted for in the analysis through statistical modeling. Thus, a more sophisticated approach to estimating posterior distributions for the treatment effect may be required, such as using Markov chain Monte Carlo methods [4].

In summary, we recommend a Bayesian approach for the analysis of trials with a small sample size, as it will give results that will help clinicians make treatment decisions. The inclusion of a prior is only likely to be acceptable if it is based on believable data and, although using noninformative priors may be preferable, it is not necessarily a straightforward option. With small sample sizes, it may be necessary to estimate the treatment effect within a statistical model in order to adjust for the likely imbalances in prognostic factors between the treatment groups.

Acknowledgements

The authors would like to thank K Abrams, Professor of Medical Statistics at University of Leicester, for mentoring in Bayesian methodology and to A Mander at the MRC Biostatistics Unit in Cambridge for first suggesting the terminology 'likelihood-based Bayesian analysis'. The authors would also like to thank M Pritchard who piloted the work on Merkel cell carcinoma as his project for MSc Clinical Oncology at the University of Birmingham.

Financial & competing interests disclosure

L Billingham and K Malottki are supported by a grant from the Medical Research Council (grant number G0800808). L Billingham is also supported by Cancer Research UK. This research is partially supported by the European Network for Cancer Research in Children and Adolescents (FP7-HEALTH-F2-2011, contract number 261474). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 European Medicines Agency Committee for Medicinal Products for Human Use. Guideline on clinical trials in small populations. *CHMP/EWP/83561/2005* (2005).

2 Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. *J. Clin. Epidemiol.* 64, 1085–1094 (2011).

3 Berry DA. Bayesian clinical trials. *Nat. Rev. Drug Discov.* 5, 27–36 (2006).

4 Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* John Wiley and Sons Ltd, Chichester, UK (2004).

5 Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *BMJ* 311, 1621–1625 (1995).

6 Matthews JNS. Small clinical trials: are they all bad. *Stat. Med.* 14, 115–126 (1995).

7 Tan S-B, Dear KBG, Bruzzi P, Machin D. Strategy for randomised clinical trials in rare cancers. *BMJ* 327, 47–49 (2003).

8 Tan S-B, Dear KBG, Bruzzi P, Machin D. Towards a strategy for randomised clinical trials in rare cancers: an example in childhood S-PNET. *BMJ* 327, 47 (2003).

9 Tan S-B, Wee J, Wong H-B, Machin D. Can external and subjective information ever be used to reduce the size of randomised controlled trials? *Contemp. Clin. Trials* 29, 211–219 (2008).

10 Burton PR. Helping doctors to draw appropriate inferences from the analysis of medical studies. *Stat. Med.* 13, 1699–1713 (1994).

11 Hughes MD. Reporting Bayesian analyses of clinical trials. *Stat. Med.* 12, 1651–1663 (1993).