Computational Statistics and Data Analysis 56 (2012) 1221–1231

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Sequential methods for pharmacogenetic studies

Susan Todd [a,*], M. Fazil Baksh [a], John Whitehead [b]

[a] Department of Mathematics and Statistics, University of Reading, Philip Lyle Building, Reading, RG6 6BX, UK

[b] Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Fylde College, Lancaster University, LA1 4YF, UK

ARTICLE INFO

Article history:
Available online 2 March 2011

Keywords:
Association studies
Candidate gene
Genome-wide
Group-sequential
Interim analyses

ABSTRACT

A study or experiment can be described as sequential if its design includes one or more interim analyses at which it is possible to stop the study, having reached a definitive conclusion concerning the primary question of interest. The potential of the sequential study to terminate earlier than the equivalent fixed sample size study means that, typically, there are ethical and economic advantages to be gained from using a sequential design. These advantages have secured a place for the methodology in the conduct of many clinical trials of novel therapies. Recently, there has been increasing interest in pharmacogenetics: the study of how DNA variation in the human genome affects the safety and efficacy of drugs. The potential for using sequential methodology in pharmacogenetic studies is considered and the conduct of candidate gene association studies, family-based designs and genome-wide association studies within the sequential setting is explored. The objective is to provide a unified framework for the conduct of these types of studies as sequential designs and hence allow experimenters to consider using sequential methodology in their future pharmacogenetic studies.

1. Introduction

A sequential study is an experiment in which the design includes one or more interim analyses that could lead to a definitive answer concerning the primary question of interest. Such a study is different from its fixed sample size counterpart in that the sample size is not calculated in advance. Instead, a stopping rule is defined which determines when the study is completed. Such studies offer practical, economic and ethical advantages through avoiding continuation of the study in the face of mounting evidence favouring a particular hypothesis. Sequential methods have been successfully implemented and have demonstrated benefits for both patients and trial sponsors in the various phases of traditional clinical development (Jennison and Turnbull, 2000; Whitehead, 1997).

Pharmacogenetics is the study of how genetic variation determines response and side-effects of therapeutic agents and the identification and design of novel drug targets. It is a rapidly developing field, see for example Brockmöller and Tzvetkov (2008), as the search continues to identify successful therapies and the best target patient populations to receive them. Recent interest in pharmacogenetics in the fields of cancer (Huang and Ratain, 2009), antiretroviral medicines (Hughes et al., 2008) and anti-epileptic drugs (Baksh and Kelly, 2008; Löscher et al., 2009) illustrates the important impact that the results of such work is likely to have on new therapeutic strategies and future policy-making decisions.

Many traditionally designed pharmacogenetic studies are not sufficiently powerful for reliable conclusions to be drawn or, as in studies of rare genotype groups and small effect sizes, require prohibitively large samples (Kirchheiner et al., 2005). Under such circumstances, it has been shown (Baksh et al., 2006; Cui et al., 2009; Dreyfus et al., 2001; Shuster et al., 2002; van

* Corresponding author. Tel.: +44 118 378 8917; fax: +44 118 378 8032.
E-mail address: s.c.todd@reading.ac.uk (S. Todd).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.02.019

der Tweel and van Noord, 2000) that use of a sequential design can be beneficial, as sequential designs are likely to use fewer observations than fixed sample designs without compromising the reliability of the conclusions. The objective of this paper is to describe and present a unified framework for the conduct of pharmacogenetic studies in clinical drug development and in post-marketing studies, within a sequential setting. The methods discussed in this paper have largely been motivated by sequential methods for genetic and epidemiological studies described elsewhere and cover candidate gene, genome-wide and family-based studies. In Section 2 an introduction to sequential methodology is given and the idea of a unified framework is introduced. Sections 3–5 then build upon this, describing methodology within the stated framework, for candidate gene association studies, family-based studies and genome-wide studies respectively. Where relevant, literature by other authors is cited and the links between their various approaches and that presented in this paper are established. Two worked examples are presented. Our aim is to provide the knowledge to allow future experimenters to consider planning their pharmacogenetic studies sequentially, in the interests of promoting more efficient research.

2. An introduction to sequential methodology

In sequential analysis, individual observations or groups of observations are analysed at interim looks to determine whether a study should be stopped. Traditionally, the term 'sequential design' has referred to looking at data after every observation while 'group-sequential design' specifies the situation where new data on a group of patients are available for each interim analysis. Modern sequential designs as described by, for example Jennison and Turnbull (2000) or Whitehead (1997) can be easily implemented in both of the above cases and in this paper, we will use the term 'sequential study' or 'sequential design' to apply to either. In order to set up a sequential study, consider a trial where the hypothesis of interest is expressed in terms of a single parameter θ and take the null hypothesis to be $H_0 : \theta = 0$. A pair of test statistics is calculated from the data available at each interim analysis of the sequential test procedure. These statistics are used to assess progress of the study and are compared with an appropriate stopping rule, used to determine whether the trial should stop or not, which is specified in advance of the trial. The procedure can be formulated so that one of these statistics, which we will denote by Z, measures the cumulative strength of evidence against the null hypothesis, H_0. The second statistic, which we will denote by V, represents the information about θ that is currently available and is closely related to sample size. A plot of Z against V forms a sample path which represents the progress of the study. This characterisation and notation, which underlies the basis of our unified approach, follows Whitehead (1997): note that the variance of Z is V rather than 1 as it would be in an alternative standardised representation. The route to implementation of a sequential procedure is: first, definition of the parameter of interest, θ, then determination of the appropriate forms for the test statistics Z and V, and finally specification of a suitable stopping rule depending on the objectives of the study. Many other representations of sequential tests involve plotting or following statistics which are exact or approximate functions of Z and V, so that they could be transformed to the structure studied here.

Once basic definitions of the parameter of interest and the test statistics have been established, attention can turn to the selection of an appropriate stopping rule. The choice is usually made on the basis of the circumstances under which small samples are desirable and will be influenced by economic and ethical considerations. Although authors have suggested slightly different approaches (for example the boundaries approach described by Whitehead, 1997 or the α-spending function approach Lan and DeMets, 1983) and indeed, the use of particular designs for particular types of study, it is important to realise that it is possible to "mix and match". The statistics Z and V relating to a particular study type can be used in conjunction with any of the designs available in the literature according to the specific requirements of the study.

The sequential framework discussed in this paper assumes normality of the test statistic Z, together with independence of the increments $(Z_i – Z_{i-1})$ between the $(i-1)$th and ith interim analysis. Any sequential procedure devised utilising these assumptions can therefore be used to conduct a genetic study. For such procedures, it is possible to evaluate properties such as the mean, median and 90th percentile of the amount of information, V, needed to complete the study. Translation is then made from V to sample size in order to explore the properties of a design under different values of θ.

In each of the following three sections we focus on the three pharmacogenetic study designs, as highlighted above. We briefly review the literature to consider where sequential approaches have been used in the past. This is followed by descriptions of the implementation of the pharmacogenetic designs in the sequential context, bringing methodology together under the common framework, considering parameterisation of the question of interest, derivation of appropriate test statistics and choice of stopping boundaries.

3. Candidate gene association studies

3.1. Case-control designs

3.1.1. Introduction and literature

The case-control study design is one of the most common designs used to assess the pharmacogenetic effects of candidate genes. The two groups, cases and controls, are distinguished by two different clinical outcomes, for example whether or not a particular type of adverse event, or a non-response, has occurred. Each group is then genotyped for a candidate gene suspected to be related to outcome. In a nested study, cases and controls are identified from well-characterised, existing controlled clinical cohorts.