

Original citation:

Hee, Siew Wan, Hamborg, Thomas, Day, S., Madan, Jason, Miller, F., Posch, M., Zohar, S. and Stallard, Nigel. (2015) Decision-theoretic designs for small trials and pilot studies: a review. *Statistical Methods in Medical Research*. pp. 1-17.

[doi:10.1177/0962280215588245](https://doi.org/10.1177/0962280215588245)

Permanent WRAP url:

<http://wrap.warwick.ac.uk/66050>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

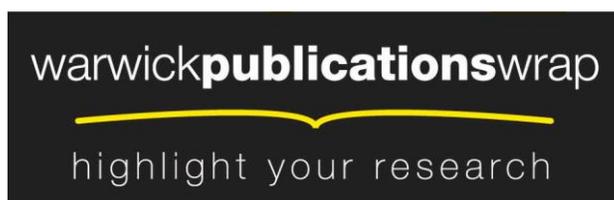
Publisher's statement:

© 2015 The Author(s)/Sage Publications Ltd. [doi:10.1177/0962280215588245](https://doi.org/10.1177/0962280215588245)

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



<http://go.warwick.ac.uk/lib-publications>

Decision theoretic designs for small trials and pilot studies: a review

Siew Wan Hee¹, Thomas Hamborg¹, Simon Day², Jason Madan¹, Frank Miller³, Martin Posch⁴,
Sarah Zohar⁵ and Nigel Stallard¹.

Abstract

Pilot studies and other small clinical trials are often conducted but serve a variety of purposes and there is little consensus on their design. One paradigm that has been suggested for the design of such studies is Bayesian decision theory. In this article we review the literature with the aim of summarising current methodological developments in this area. We find that decision theoretic methods have been applied to the design of small clinical trials in a number of areas. We divide our discussion of published methods into those for trials conducted in a single stage, those for multi-stage trials in which decisions are made through the course of the trial at a number of interim analyses, and those that attempt to design a series of clinical trials or a drug development programme. In all three cases a number of methods have been proposed, depending on the decision maker's perspective being considered and the details of utility functions that are used to construct the optimal design.

Keywords

Bayesian decision theory, optimal clinical trial design, phase II clinical trials, utility functions

¹Division of Health Sciences, Warwick Medical School, University of Warwick

²Clinical Trials Consulting and Training Limited, Buckingham, UK.

³Department of Statistics, Stockholm University

⁴Section of Medical Statistics, CeMSIIS, Medical University of Vienna, Austria

⁵INSERM, U1138, team 22, Centre de Recherche des Cordeliers, Université Paris 5, Université Paris 6, Paris, France

Corresponding author

Dr Siew Wan Hee, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK. E-mail: s.w.hee@warwick.ac.uk.

Decision theoretic designs for small trials and pilot studies: a review

1. Introduction

There is little consensus on the way in which pilot studies and other small exploratory clinical trials should be designed, with a relatively wide range of approaches proposed.¹⁻⁴ In part, this reflects the range of objectives for such studies. Pilot studies are usually designed to explore and evaluate the efficacy and safety of a new/experimental treatment or new combination/regime of treatments, perhaps to provide some evidence of response in order to justify the financial input required for larger-scale studies, though they may also address specific additional or alternative research questions. Generally, the sample size required in pilot studies is quite small, as they commonly precede a larger definitive clinical trial. The terminology of such small exploratory clinical studies reflects the range of objectives, with small trials that are conducted prior to a major definitive study referred to as ‘pilot studies’, ‘feasibility studies’ or – if the following study is a ‘phase III’ trial – ‘phase II clinical trials’. While a phase II clinical trial can be in some cases relatively large, it usually has key objectives to decide whether or not and how to conduct the following phase III trial(s). The different terminology depends sometimes on, often fairly subtle, differences in the aim of the study, but also on the setting in which the study is being conducted, with different names being used for trials with essentially the same purpose conducted by pharmaceutical companies and in the public sector, for example.

The variety in design approaches in pilot studies and other small early phase trials is in contrast to the setting of confirmatory or phase III clinical trials where almost all trials are designed using a frequentist paradigm so as to control probabilities of a type I error, corresponding in this setting to claiming an experimental treatment is better than the control treatment when it is not, and of a type II error, that is failing to claim that the experimental treatment is better when it is by some specified magnitude. The type I error rate is usually, although only by convention, set at a (two-sided) level of 0.05 and the type II error rate is conventionally set at 0.10 or 0.20, corresponding to a power of 0.9 or 0.8, respectively.^{5,6}

The frequentist approach with conventionally used error rates for a typical effect size may lead to the relatively large sample sizes associated with confirmatory studies, so may not be appropriate when the sample size is smaller. This is the case in pilot studies and early phase trials. Small trials may also be unavoidable, for example, in the setting of trials in small population groups such as patients suffering from a rare disease, patient groups where patient recruitment is difficult such as children and other vulnerable populations or in a specifically targeted subpopulation. In these settings, then, either the

frequentist approach can be applied with error rates relaxed, or some other method must be used to design the trial.⁷⁻¹⁰

A number of novel approaches have been proposed for pilot studies and other small clinical trials. As the sample sizes are small, often with recruitment occurring relatively slowly, sometimes in a single centre, multi-stage designs with decisions made at one or more interim analyses are often an attractive option, as are multi-arm trials, widening the range of design choices further. Alternatives to the frequentist paradigm that have been proposed for the design of such small clinical trials include the Bayesian approach and the decision theoretic approach in which, as described in the next section, the consequences of decisions are explicitly modelled. The latter might seem particularly appropriate in pilot studies and early phase clinical trials as the outcome from such trials is often a relatively simple decision that is within the control of the clinical team conducting the trial, such as the decision whether or not to conduct further clinical research in the area, sometimes called ‘Go/No-Go’ decisions.

The aim of this article is to review methods for the design of small trials and pilot studies where the primary aim is to explore and evaluate treatment efficacy based on the Bayesian decision theoretic framework. In order to provide as comprehensive as possible a review of the current literature in this area, we used systematic reviewing methodology to identify relevant published work in the area. In addition, human pharmacology studies that aim to assess toxicity, to explore drug metabolism and drug interactions, or to describe the pharmacokinetics and pharmacodynamics are usually designed as small studies. These studies are sometimes known as ‘phase I’ trials. Although the sizes of these studies tend to be small, their objectives are not to explore and evaluate the efficacy of a new/experimental or new combination/regime of treatments. As such, these studies are excluded from the review of this paper.

Pilot or feasibility studies may also be used to test practical aspects intended to be used in a later study such as drug supply, acceptability of randomisation or visit schedules and so on. Such studies, which may closely resemble a larger study conducted in miniature are, however, outside the scope of this paper.

Following this introduction, the next section of the article gives a very brief overview of the decision theoretic approach as it may be applied to clinical trial design. The third, and most substantial, section of the article then describes the review method used and gives details of publications found. Papers are classified into a number of groups according to the specific type of design being considered, the types of utility function proposed and the perspective of the decision maker (commercial, regulatory/societal or patient). The aim is to allow the reader to rapidly identify key references in each

particular area. The paper ends with a brief discussion of the scope of the review, suitability of the decision theoretic paradigm for pilot study design and suggestion of areas where further research work might be most appropriate.

2. The decision theoretic approach to clinical trial design

Decision theory is a statistical technique by which the problem of decision-making under uncertainty may be formalized. The method enables an optimal decision to be made between a number of possible actions on the basis of the consequences of each action under all possible scenarios.¹¹⁻¹³

In the setting of a clinical trial, we may wish to decide between a number of possible design options. Prior to the start of the trial, this might entail a choice of the clinical trial sample size. During the trial at an interim analysis, a decision might be taken as to whether or not to terminate the trial or to modify the trial conduct in some way. At the end of the trial this might correspond to a decision of whether or not to proceed with further trials, or in a multi-arm study, to choose an experimental treatment for further evaluation.

Denote by Y the vector of responses from n patients in a clinical trial, with Y assumed to follow a distribution with density function of known form with some unknown parameter(s), θ . Having observed $Y = (y_1, y_2, \dots, y_n)$, a decision, to choose one from a set of possible actions, $\mathcal{A} = \{a_1, a_2, \dots\}$, is made. Suppose that the consequences associated with each possible action can be expressed by some loss or gain. As the consequences may well depend on the true unknown state of nature, the loss or gain will be a function of parameter θ . This function is called a loss, gain, or utility function. Therefore, the utility function for action a may be written as $U_a(\theta)$.

Uncertainty regarding the parameter θ may be expressed in terms of a Bayesian prior distribution, or, following observation of some data, by the corresponding posterior distribution. The expected utility from taking action a is then the expected value $E(U_a(\theta))$, the expectation being taken over this prior (or posterior) distribution. Comparing expected utility values for different actions enables the optimal action to be determined.

The utility functions should express the values of the consequences of possible actions from the perspective of the decision-maker. These could be monetary loss or reward, which is measurable on an existing scale, or could also express consequences that have no immediately obvious numerical scale of measurement, such as treatment success (patient experienced a positive response) or

treatment satisfaction. In the latter cases it can be very difficult to assign the numerical values required to specify the utility function to the qualitative values, as considered in the discussion section below.

Similarly, the prior distribution should reflect prior belief regarding the parameters of the distribution of the responses. The source of information for the prior distribution may be obtained from data from previous similar trials, elicitation of expert opinion or, as a conjugate prior, for computational convenience.

Once utility functions have been specified for all possible actions, the optimal trial design can often be obtained by working in a reverse time order using a method known as ‘dynamic programming’ or ‘backward induction’. For example, suppose we wish to obtain the optimal sample size for a single stage trial with a number of possible actions available at the end of the trial. First, all possible actions at the end of the trial are considered and, for given possible observed data, Y , their corresponding expected posterior utilities computed and compared to obtain the optimal action a given Y . Second, since at the planning stage Y is unknown, the expected utility (assuming after the trial the optimal action is chosen), is obtained from the distribution of Y given n and θ . Finally, because also θ is unknown, the expectation of this expected utility is taken over the prior distribution for θ . Thus, overall the expected utility is computed, taking the expectation over the prior predictive distribution for Y given n . This gives the prior expected utility from a trial of sample size n if the optimal action is always taken at the end of the trial. This is a function of n alone, so maximizing this over n gives the optimal sample size, which is thus obtained by finding

$$\operatorname{argmax}_n \left\{ \int_y \max_a \left\{ \int U_a(\theta) p(\theta|y, n) d\theta \right\} f(y|n) dy \right\}$$

where $f(y|n)$ is the density function for y before sampling and with only prior knowledge of θ , that is the prior predictive distribution of y .¹⁴ Note that for discrete responses the expectation may be the summation for all possible values of y . This method can be extended to multi-stage trials with optimal decisions at the end of a given stage obtained by considering the consequences in terms of decisions taken later in the trial.

Although relatively easily described as above, the application of decision theory in a clinical trial setting can present a number of challenges: in specification of appropriate prior distributions and, perhaps more especially, in specification of utility functions. When considering decisions made during or prior to the start of a trial, calculations of how the utility functions should incorporate the impact of data as yet unobserved should also be taken into consideration. The required computation can also be challenging, particularly in the case of multi-stage trials.^{15, 16}

3. Review of decision theoretic approaches to pilot studies and small clinical trials

3.1 Literature review search strategy and results

We undertook a systematic search in the electronic database Scopus for research articles published up to 10 October 2014 to identify work on decision-theoretic methods for the design of pilot studies and small clinical trials.¹⁷ The exact search terms used and number of hits from each of these terms are shown in Appendix A.1. The searches were not limited by subject area. A total of 8751 articles were identified of which 425 were duplicates, leading to a total of 8326 unique articles. We also identified a further 10 articles via screening of references in relevant articles. Figure 1 shows the flow diagram of articles identified, excluded and included in this review. Articles were screened by title, abstract or whole paper for relevance to this review, leading to final inclusion of 67 articles which are discussed in detail below^{14, 18-83}. Details of the 67 articles are given in Table 1.

Twenty seven articles¹⁸⁻⁴⁴ specifically described methods in pilot or phase II settings. The others^{14, 45-83} did not describe methods specific to small clinical trials or pilot studies, but would nevertheless be appropriate in this setting.

A study of the articles identified indicated that the design problems considered fell into three broad categories, with different approaches used for each. The simplest type of design considered was that for clinical trials conducted in a single stage so that decisions considered are those taken at the end of the trial and those taken regarding trial design prior to the start of the trial essentially using the method outlined above. The second type of design considered was that for multi-stage trials, so that decisions taken through the course of the trial, at a number of interim analyses, are also considered. The third type of design considered was that concerning multi-arm trials or the simultaneous optimal design of a series of clinical trials. Within each of these three design types there was a variety of approaches corresponding to the viewpoint of the decision-maker and the complexity of the utility functions considered, ranging from approaches in which a relatively simple utility function aims to reflect the number of patients successfully treated, to more complex utility functions based on a detailed elicitation of the costs and consequences of a range of possible outcomes from the perspective of a particular decision-maker. Table 1 shows the identified articles classified by types of design, utility function and perspective of the decision-maker. The literature in each of the areas identified is described in the following subsections. One paper⁴² fell outside of the three categories just described, being concerned with the design of enrichment studies. This paper is also discussed below.

3.2 Single-stage designs

The simplest type of study design considered is the single-stage design in which no analysis of the data is conducted until the study is completed. For a trial comparing an experimental treatment with a control (a two-arm trial) or a historical control (a single-arm trial), the statistical design choices are relatively limited; the main one being the choice of sample size. Authors whose works considered this type of design include Brunier and Whitehead,¹⁹ Chen and Beckman,²⁰ Chen et al.,²¹ Claxton and Posnett,⁵² Claxton and Thompson,⁵³ Eckermann and Willan,⁵⁴ Gittins and Pezeshk,⁵⁵⁻⁵⁷ Halpern et al.,⁵⁸ Hornberger and Eghtesady,⁶⁰ Kikuchi and Gittins,^{63, 64} Kikuchi et al.,⁶⁵ Lindley,¹⁴ Maroufy et al.,⁶⁸ Patel and Ankolekar,⁷¹ Pezeshk and Gittins,⁷³ Pezeshk et al.,^{74, 75} Staquet and Sylvester,³⁸ Sylvester,³⁹ Sylvester and Staquet,⁴⁰ Willan,⁷⁹ Willan and Eckermann,^{80, 81} and Willan and Pinto.⁸³

At the end of such a trial a decision is made from a possible set of actions which typically consists of whether or not to accept the experimental treatment for further study. The utility function may be written simply as a function of the cost of sampling, which is independent of θ , the benefit of treating patients with the experimental treatment and the cost of making an incorrect decision.^{14, 20, 38-40} The cost of sampling may reflect the costs, interpreted broadly, of recruiting, monitoring or treating patients in the trial, and can be expressed as a cost per patient, in some cases expressed in monetary units. If the experimental treatment is accepted for further study and it is indeed effective then some expected number of future patients will benefit from it. In considering the cost of making an incorrect decision, (that is accepting an ineffective experimental treatment or rejecting an effective experimental treatment) one may assign an absolute value to each of these consequences but this is not as easy as assigning monetary rewards or loss. As the utility function expresses the consequences of actions in relation to each other, one approach is to order the actions in terms of seriousness and to assign the less serious consequence some arbitrary utility value, say 1, with the other more serious consequence assigned some multiple of this. Finally, both the benefit of health outcome of future patients and the cost of incorrect decisions are rescaled to monetary terms so that these can be combined in the same utility function. Alternatively, the cost of sampling may be rescaled to a health outcome. For example, in the health economic 'Value of Information' approach, costs and benefits in terms of quality of life are often expressed in Quality-adjusted Life Years. As the benefit and cost are in relation to each other, some authors have proposed a model that is a function of the benefit-cost ratio.²⁰ Alternatively, the utility function may encompass a more realistic setting with health outcome and economic criteria such as the profit from treating future patients, market share, the costs incurred from conducting the trial and the costs of sampling explicitly included.^{19, 21, 52-58, 60, 63-65, 68, 71, 73-75, 79-81, 83}

Designs that considered simple utility functions did not generally specify the perspective of the decision-maker,^{14, 38-40} whereas more realistic utility functions are often explicitly based on a commercial,^{21, 55-58, 65, 68, 71, 73, 75, 79-81} regulatory,^{54, 56, 57, 73, 74, 81} or societal perspective.^{52, 53, 55, 60, 63, 64, 83} Exceptions are the work of Chen and Beckman²⁰ where the utility function is a simple function based on a commercial perspective that controls for type I and II error rates under the constraint of limited sample size, and Brunier and Whitehead¹⁹ where the utility function is a more realistic one that incorporates costs that may be incurred during the trial but there was no specification of the perspective of the decision-maker.

Possible actions for designs based on a societal perspective include whether to delay the decision making and start a new trial, to adopt the experimental treatment (granting licence or reimbursement of costs of the treatment in general clinical use) and start a new trial to gather more information or to adopt the experimental treatment without starting a new trial (that is, requiring no further information).^{54, 74, 81}

3.3 Multi-stage designs

A more complex decision process arises in trials that are conducted in a number of stages with interim analyses conducted at the end of each. The majority of clinical trials design methods based on decision theory have considered this setting, usually with the possible actions at the end of each stage taken to be those corresponding to stopping the trial for futility, stopping with a positive result or continuing to further stage(s). Specific articles include Banerjee and Tsiatis,¹⁸ Berry and Ho,⁴⁵ Chen and Willan,⁴⁶ Chen and Smith,²² Cheng and Berry,⁴⁷ Cheng and Shen,^{48, 49} Cheng et al.,⁵⁰ Chernoff and Petkau,⁵¹ Ding et al.,²³ Heitjan et al.,⁵⁹ Jennison and Turnbull,⁶¹ Jiang et al.,⁶² Jung et al.,²⁵ Lewis and Berry,⁶⁶ Lewis et al.,⁶⁷ Mehta and Patel,⁶⁹ Nixon et al.,²⁷ Orawo and Christen,⁷⁰ Palmer,²⁹ Rossell et al.,³⁰ Stallard,^{31, 32} Stallard et al.,³⁷ Wang,⁷⁶ Wathen and Christen,⁷⁷ Wathen and Thall,⁷⁸ Willan and Kowgier,⁸² and Zhao et al.,⁴³ and Zhao and Woodworth.⁴⁴

A multi-stage trial may be designed based on consideration of a fixed and known maximum number of patients that can be included in the trial. In rare diseases settings, one may be able to estimate the number of cases eligible for clinical trials relatively easily, or it can be tied to budget allocation. In such scenarios, at the final stage when all patients have been recruited there are two terminal actions to choose; stop and accept the treatment for further study or stop and reject the treatment from further study.

Some designs are aimed to optimize the patient allocation for a fixed and known number of future patients, known as the 'patient horizon', N , which may be estimated via the incidence rate for the

disease of concern and an assumption regarding the life of any potential new treatment. The aim is to choose n , the number of patients that are allocated to the pilot or phase II trial assuming the remaining $N - n$ are either included in the larger phase III trial or subsequently receive the recommended treatment.^{47, 50, 51, 76, 82} The expected utility for each terminal action (stop and accept treatment or stop and reject treatment) is given directly by the utility function specified, whereas the expected utility from continuing to the next stage depends on the observations and decisions made at a future stage. Thus, to estimate the expected utility function for each action at each stage we begin by considering the terminal actions at the ultimate stage supposing that responses from all N patients have been observed. Having estimated the maximized utility function we can go one step back and estimate the expected utility function for all possible actions prior to the penultimate stage. Following this manner of iteration the expected utility function for each action at the first stage is computed, extending the method outlined above to a full backward induction approach.

The backward induction computation becomes very intensive as the number of stages increases. Orawo and Christen, and Wathen and Christen have considered using an approximation method rather than computing the exact value.^{70, 77} For designs that do not assume a fixed known patient horizon, the optimal sequential design can sometimes be computed by forward simulation and constrained backward induction.^{23, 30, 43, 78} Heitjan et al. consider a two-stage design,⁵⁹ reducing the computational burden considerably, and use direct numerical optimization rather than the backward induction approach to obtain optimal designs in this case, while Jung et al. adopt a similar approach for two-stage single-arm cancer trials,²⁵ comparing their approach with the commonly used design due to Simon.⁸⁴

Just as we had a range of single stage designs considered above, authors have taken a number of approaches to the design of multi-stage trials. Some authors have considered simple utility functions (costs of making incorrect decisions and cost of sampling) in their designs.^{18, 25, 29, 43, 44, 47-51, 59, 61, 62, 66, 67, 76-78} Cheng and Shen⁴⁹ related the utility function parameters to frequentist error rates, while Nixon et al.²⁷ related it to the expected prior probability of success which is sometimes known as assurance, a term introduced by O'Hagan and Stevens.⁸⁵ Frequentist error rates are also considered by Jennison and Turnbull who use backward induction to obtain group sequential designs that are optimal in that the expected sample size is minimized subject to the error rate requirements.⁶¹ Others have considered more realistic utility functions.^{22, 23, 30-32, 37, 43, 45, 46, 69, 70, 82} In this latter case the utility function is usually constructed from a commercial perspective.

3.4. Enrichment designs

Trippa et al. proposed a decision-theoretic approach to an enrichment design.⁴² In this two-stage design, all patients enrolled to stage one receive the same experimental treatment. The data from these patients are then used to optimally identify the population of patients to be used in the main, second, stage of the trial, in which patients are randomized to receive either the experimental or the control treatment. In the design proposed by Trippa et al. the utility function encompasses the benefit that will be received by future patients if the experimental treatment is recommended for further study in a phase III setting in the population identified, the costs incurred for conducting the phase II and III trials, and the duration of treatment in stage one which influences the population of patients for whom treatment is successful that will be used in the second stage.

Although represented by a single paper in our review, the area of enrichment designs is one of considerable recent statistical interest (see, for example, Graf et al.,⁸⁶ Simon and Simon,⁸⁷ and Wang et al.⁸⁸), suggesting that this is an area in which new work on decision-theoretic approaches might be anticipated.

3.5. Designs for multi-arm trials, programmes of studies or a series of trials

Some articles extend the multi-stage designs for single-arm trials or two-arm comparative trials to seek optimal multi-arm designs. Such a problem is considered by Chen and Beckman,²⁰ Lai et al.,²⁶ Palmer,²⁹ Patel and Ankolekar,⁷¹ Patel et al.,⁷² Stallard et al.,³⁵ and Thall et al.⁴¹ This introduces, in addition to possible actions corresponding to stopping or continuing the trial, the option of dropping one or more treatment arms at an interim analysis, so that the decision process can become increasingly complex.

In multi-arm trials, J doses of the experimental treatment or J different experimental treatments may be tested against the standard treatment or placebo. At each interim analysis, a decision is made to drop inferior dose(s) or treatment(s), and to continue recruitment to the remaining dose(s) or treatment(s). At the final stage, a decision is made to either recommend the superior dose or experimental treatment for further study in a phase III trial setting, or to reject it from further study. Note that this type of study is different from dose-finding trials. The J doses are assumed to have been recommended from dose-finding trials.

In the designs considered above, patients are generally considered in groups, with decisions made at interim analyses once the data from each group of patients has been observed. If patients are considered one at a time, the problem becomes one of optimally allocating treatments to each patient. Such an approach is considered in a two-arm study by, e.g., Jiang et al.⁶² This problem is closely related to the multi-armed bandit problem. In some settings the different treatments being compared in

a multi-arm design may be different doses of the same drug. In this case a parametric dose-response model may be assumed. These two settings are considered briefly in the discussion section below.

The objective of the designs considered by Patel and Ankolekar⁷¹, and Patel et al.,⁷² is to maximize the expected profit from a portfolio of treatments while incurring the costs of running the trials within the given budget. An optimal size is obtained for each treatment (each trial) and the trials may run concurrently.

Some articles use decision theory methods to design not a single study but a series of studies, which may themselves employ single stage or multi-stage designs. The problem of decision-making at the end of one trial in a series of potential trials is rather like that at an interim analysis in a multi-stage trial, so that the methods often build on those described above. In this setting actions corresponding to moving on from one trial to another need to be considered. Articles considering designs for this setting include Chen and Beckman,²⁰ Hee and Stallard,²⁴ Pallay,²⁸ Stallard,^{33, 34} and Stallard and Thall.³⁶ The trials, usually with one experimental treatment per trial (either a single-arm design or two-arm comparison with a control), run sequentially, so that a decision made in one trial can affect possible future trials. In a series of single stage trials, suppose n is the sample size of a trial and upon observing all the responses a decision is made either to accept the experimental treatment for further study or to reject it from further study and start a new trial with a different treatment. In a series of multi-stage trials, at each interim stage a decision is made whether to continue recruitment to the current trial or to terminate the trial and recommend the experimental treatment for further study or to initiate a new trial with a different experimental treatment. Some designs may also have a possibility of terminating the current trial and abandoning the whole development plan.²⁴

The full backward induction approach can be very challenging in this case, so authors have generally either considered small sample sizes,³⁶ or sought simpler algorithms or asymptotic results to give approximately optimal designs.^{47, 50}

The expected utility for a series of sequential trials may also be computed via a backward induction algorithm similar to that described above. For a series of multi-stage trials the backward induction is used within each trial as well as for the series of sequential trials. Almost all assumed a commercial perspective with realistic utility functions.^{24, 28, 33, 36} Both Chen and Beckman, and Stallard, on the other hand assumed a commercial perspective with simple utility functions.^{20, 34}

4. Discussion

The aim of this article has been to review the literature on methods for pilot studies and small clinical trials that are based on the use of Bayesian decision theory. Methods have been published for single stage and multi-stage clinical trials as well as for multi-arm trials or series of trials, with utility functions based on a number of different decision makers' perspectives. Most methods have focussed on a decision regarding the sample size of the trial, though in general other features of the design could be chosen in a similar way. Specific examples that have been considered include dropping of arms in a multi-arm trial and selection of the population in the enrichment design.

It is inevitable that when writing an article such as this decisions must be made regarding the scope of the review. Within the limit of Bayesian decision-theoretic methods, our intention has been to keep the scope fairly wide, including discussion of methods for any clinical trial design that might be appropriate for a small trial or pilot study with an efficacy endpoint. One exclusion has been methods for phase I or dose-finding studies, where the main concern is usually a safety or toxicity endpoint. Although Bayesian and decision-theoretic methods are relatively common in this setting, the different endpoint, use of sequential designs with very small groups, often making decisions after each subject, and incorporation of dose-response information (so that data from one arm can lead to inference regarding other arms) mean that the methods proposed are rather different to those we have considered, and are less suited to other small trials. Readers interested in this area are directed to the work by Cheung⁸⁹ and Simes⁹⁰. One of the few published applications of the decision-theoretic methodology is in the phase I oncology setting⁹¹. Our choice of search terms also excluded literature on the multi-arm bandit problem, identifying one paper⁷⁶ applying this methodology specifically to clinical trial design with the intention of optimally allocating patients one at a time to treatments in a multi-arm study. There is a relatively large body of literature on this problem in applied probability journals which, although considered from a more generally viewpoint might be relevant to clinical trials of this type.⁹²⁻⁹⁴

A challenge, in any Bayesian methodology, is the specification of a prior distribution. Most of the papers identified in the review used conjugate prior distributions to facilitate mathematical derivation.^{14, 22-24, 26-28, 30-37, 41-51, 54-58, 60, 62-70, 73-83} In most cases this involved using a beta prior distribution for a Bernoulli distribution, or in some cases taking a two-point prior corresponding to an experimental treatment that is either effective or ineffective.^{18, 20, 21, 25, 29, 38-40, 52, 53, 59, 61, 72} In some cases, the prior distribution may be for a vector of unknown parameters. Some examples are normal distribution with both unknown mean and variance^{57, 63, 65} or a time-to-event endpoint where the hazard function is modelled with a 3-parameter generalized gamma and the unknown priors follow a gamma and inverse gamma distributions⁷⁸. Authors whose works consider more than one endpoint, for example, Bernoulli efficacy and Bernoulli toxicity, assume a Dirichlet distribution^{22, 37, 75}, or for time-to-event and Bernoulli toxicity endpoints where the unknown parameters follow a bivariate

gamma (regression) distribution⁴¹ or gamma and beta distributions²⁶. Some papers used MCMC methods,^{43, 68} and one paper was based on a non-parametric approach.⁴³ When priors were specified, they were usually informative, sometimes with a number of alternative priors used and results compared. Although there is a considerable literature on elicitation of Bayesian prior distributions (see, for example, Chaloner et al.,⁹⁵ Kadane and Wolfson,⁹⁶ O'Hagan,⁹⁷ and case studies by Blanck et al.⁹⁸, and Kinnersley and Day⁹⁹), only two articles identified in our review described the use of formal methods for prior elicitation methods.^{42, 52}

As described above, one major challenge in the development and application of decision-theoretic methods in clinical trials is that of constructing utility functions that accurately reflect the consequences of possible actions. It is clear from the articles discussed above that approaches to this challenge have varied. Some researchers have focussed on monetary costs and rewards, whilst others have compared these with improvement or deterioration in health states using approaches from health economics. The utilities should reflect the preferences of consequences from the point of view of the decision-maker. This can be particularly challenging when more than one individual or group will make a decision based on the results of a clinical trial, or be otherwise affected by the results, or if the decision-maker and the trialist have different viewpoints. For example, decision-making by a societal decision-maker such as National Institute for Health and Care Excellence (NICE) in the UK may be based primarily on cost-effectiveness whereas decision-making by a pharmaceutical company may be based more on whether or not the current information is sufficient to apply for licencing for the experimental treatment. In an attempt to reconcile this challenge, Willan and Eckermann proposed a design that combined both public health service and commercial perspectives where the utility function is made up of two thresholds, namely, a maximum price of the experimental treatment acceptable to the public health service for reimbursement and a minimum price to the pharmaceutical company that does not result in a loss of investment.⁸¹

More general methods for construction of utility values based on direct consideration of consequences have been based on prioritizing preference, for example using methods first proposed by Ramsey¹⁰⁰ or methods discussed by Lindley¹⁰¹ or Emrich and Sedransk¹⁰². Although a variety of approaches have been taken, most researchers proposing single stage designs have based utility functions on a patient or societal perspective whereas commercial perspectives have been more common in development of multi-stage designs. In the description of the method above, we have taken the utility function to depend on the unknown parameters alone, as proposed by Lindley¹⁴ and Raiffa and Schlaifer¹². Most authors proposing simple utility functions have followed this approach. Some authors have proposed more complex utility functions in which the utility depends also on the observed trial data, for example with a gain if a trial indicates a significant treatment effect. In spite of the numerous approaches proposed, it seems likely that it is this difficulty with specification of an appropriate utility

function, together with a lack of familiarity, both with Bayesian methods in general and with decision-theoretic methods in particular, that is responsible for the very limited use of decision-theoretic methods in practice.

In spite of the challenges, we consider a Bayesian decision theoretic approach to be appropriate for the design of pilot studies and early phase trials given the clear role of these trials is to inform decisions regarding further future clinical research. However such trials are designed, these decisions will be made and the decision-theoretic approach formalizes this by considering the decisions and their consequences explicitly. Even when trials are designed based on other approaches, we believe that the decision-theoretic methodology is a useful tool for trialists and statisticians designing trials, enabling the properties of trial designs obtained under one paradigm to be evaluated based on another. This is, perhaps, particularly important in small trials when compromise is inevitable, as it leads to a careful consideration of the purpose of the trial and its required properties, thus ensuring that it is fit for purpose. One thing that we believe could increase the use of decision-theoretic designs is a greater familiarity and improved understanding through retrospective evaluation of such approaches.

This review has identified many decision-theoretic approaches. In any real application it is important to consider the purpose of the trial and ensure that this is reflected in the formulation of the decision problem and utility function so that the trial design proposed is appropriate to match this purpose.

Acknowledgements

This work was conducted as part of the InSPiRe (Innovative methodology for small populations research) project funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013 – 602144.

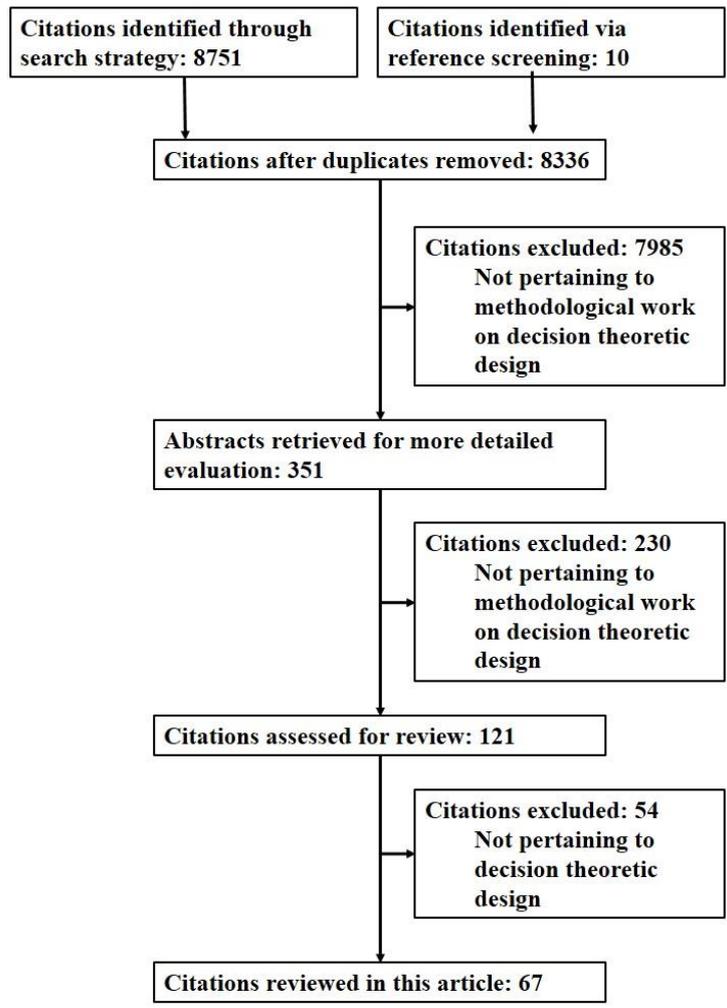


Figure 1. Flow diagram of articles identified, excluded and included for review.

Multi-stage	Single-arm	-	-	-	Banerjee and Tsiatis ¹⁸ Jung et al. ²⁵ Cheng et al. ⁵⁰ Zhao et al. ⁴³ Zhao and Woodworth ⁴⁴	-	-	Chen and Smith ²² Ding et al. ²³ Russell et al. ³⁰ Stallard ^{31, 32} Stallard et al. ³⁷	Zhao et al. ⁴³
	Two-arm	Heitjan et al. ⁵⁹	-	-	Cheng and Berry ⁴⁷ Cheng and Shen ^{48, 49} Cheng et al. ⁵⁰ Chernoff and Petkau ⁵¹ Jennison and Turnbull ⁶¹ Jiang et al. ⁶² Lewis and Berry ⁶⁶ Lewis et al. ⁶⁷ Nixon et al. ²⁷ Wang ⁷⁶ Wathen and Christen ⁷⁷ Wathen and Thall ⁷⁸	-	Willan and Kowgier ⁸²	Berry and Ho ⁴⁵ Chen and Willan ⁴⁶ Mehta and Patel ⁶⁹	Orawo and Christen ⁷⁰
Multi-arm	-	-	Chen and Beckman ²⁰ Lai et al. ²⁶	Palmer ²⁹ Stallard et al. ³⁵ Thall et al. ⁴¹	-	-	Patel and Ankolekar ⁷¹ Patel et al. ⁷²	-	
Enrichment	-	-	-	-	-	-	-	-	Trippa et al. ⁴²
Series of trials	-	Stallard ³⁴	Chen and Beckman ²⁰ Stallard ³⁴	-	-	-	Hee and Stallard ²⁴ Pallay ²⁸ Stallard ³³ Stallard and Thall ³⁶	-	

a The total number of articles in the cells exceed 67 as some described more than one design or perspective.

References

1. Lancaster GA, Dodd S and Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract.* 2004; 10: 307-312.
2. Arnold DM, Burns KEA, Adhikari NKJ, et al. The design and interpretation of pilot trials in clinical research in critical care. *Crit Care Med.* 2009; 37: S69-S74.
3. Arain M, Campbell MJ, Cooper CL and Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol.* 2010; 10: 67.
4. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol.* 2010; 10: 1.
5. Pocock SJ. *Clinical Trials: A Practical Approach.* Wiley, 1983.
6. Julious SA. *Sample Sizes for Clinical Trials.* CRC Press, 2008.
7. EMA. Guideline on clinical trials in small populations. European Medicines Agency CHMP/EWP/83561/2005, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf (2006, accessed 29 October 2014).
8. EMA. Concept paper on extrapolation of efficacy and safety in medicine development. European Medicines Agency EMA/129698/2012, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/04/WC500142358.pdf (2013, accessed 29 October 2014).
9. Institute of Medicine. *Small Clinical Trials: Issues and Challenges.* National Academy Press, <http://www.nap.edu/catalog/10078.htm> (2001, accessed 29 October 2014).
10. Casali PG, Bruzzi P, Bogaerts J and Blay J-Y. Rare Cancers Europe (RCE) methodological recommendations for clinical studies in rare cancers: a European consensus position paper. *Ann Oncol.* 2014.
11. DeGroot M. *Optimal Statistical Decisions.* Wiley, 1970.
12. Raiffa H and Schlaifer R. *Applied Statistical Decision Theory.* MIT Press, 1977.
13. Smith JQ. *Decision Analysis: A Bayesian Approach.* Chapman and Hall, 1988.
14. Lindley DV. The choice of sample size. *Journal of the Royal Statistical Society: Series D (The Statistician).* 1997; 46: 129-138.
15. Brockwell AE and Kadane JB. A gridding method for Bayesian sequential decision problems. *Journal of Computational and Graphical Statistics.* 2003; 12: 566-584.
16. Kadane J and Vlachos P. Hybrid methods for calculating optimal few-stage sequential strategies: Data monitoring for a clinical trial. *Statistics and Computing.* 2002; 12: 147-152.
17. Elsevier. Scopus Content Overview. <http://www.elsevier.com/online-tools/scopus/content-overview> (Accessed 20 October 2014).

18. Banerjee A and Tsiatis AA. Adaptive two-stage designs in phase II clinical trials. *Stat Med.* 2006; 25: 3382-3395.
19. Brunier HC and Whitehead J. Sample sizes for phase II clinical trials derived from Bayesian decision theory. *Stat Med.* 1994; 13: 2493-2502.
20. Chen C and Beckman RA. Optimal cost-effective designs of phase II proof of concept trials and associated go-no go decisions. *J Biopharm Stat.* 2009; 19: 424-436.
21. Chen C, Sun L and Li CL. Evaluation of early efficacy endpoints for proof-of-concept trials. *J Biopharm Stat.* 2013; 23: 413-424.
22. Chen Y and Smith BJ. Adaptive group sequential design for phase II clinical trials: A Bayesian decision theoretic approach. *Stat Med.* 2009; 28: 3347-3362.
23. Ding M, Rosner GL and Müller P. Bayesian optimal design for phase II screening trials. *Biometrics.* 2008; 64: 886-894.
24. Hee SW and Stallard N. Designing a series of decision-theoretic phase II trials in a small population. *Stat Med.* 2012; 31: 4337-4351.
25. Jung SH, Lee T, Kim KM and George SL. Admissible two-stage designs for phase II cancer clinical trials. *Stat Med.* 2004; 23: 561-569.
26. Lai TL, Liao OYW and Zhu RG. Adaptation in clinical development plans and adaptive clinical trial designs. *Statistics and its Interface.* 2012; 5: 431-442.
27. Nixon RM, O'Hagan A, Oakley J, et al. The rheumatoid arthritis drug development model: A case study in Bayesian clinical trial simulation. *Pharm Stat.* 2009; 8: 371-389.
28. Pally A. A decision analytic approach to a futility analysis of a phase II pharmaceutical study. *J Biopharm Stat.* 2001; 11: 209-225.
29. Palmer CR. A comparative phase II clinical trials procedure for choosing the best of three treatments. *Stat Med.* 1991; 10: 1327-1340.
30. Rossell D, Müller P and Rosner GL. Screening designs for drug development. *Biostatistics.* 2007; 8: 595-608.
31. Stallard N. Approximately optimal designs for phase II clinical studies. *J Biopharm Stat.* 1998; 8: 469-487.
32. Stallard N. Sample size determination for phase II clinical trials based on Bayesian decision theory. *Biometrics.* 1998; 54: 279-294.
33. Stallard N. Decision-theoretic designs for phase II clinical trials allowing for competing studies. *Biometrics.* 2003; 59: 402-409.
34. Stallard N. Optimal sample sizes for phase II clinical trials and pilot studies. *Stat Med.* 2012; 31: 1031-1042.
35. Stallard N, Posch M, Friede T, Koenig F and Brannath W. Optimal choice of the number of treatments to be included in a clinical trial. *Stat Med.* 2009; 28: 1321-1338.
36. Stallard N and Thall PF. Decision-theoretic designs for pre-phase II screening trials in oncology. *Biometrics.* 2001; 57: 1089-1095.

37. Stallard N, Thall PF and Whitehead J. Decision theoretic designs for phase II clinical trials with multiple outcomes. *Biometrics*. 1999; 55: 971-977.
38. Staquet M and Sylvester R. A decision theory approach to phase II clinical trials. *Biomedicine*. 1977; 26: 262-266.
39. Sylvester RJ. A Bayesian approach to the design of phase II clinical trials. *Biometrics*. 1988; 44: 823-836.
40. Sylvester RJ and Staquet MJ. Design of phase II clinical trials in cancer using decision theory. *Cancer Treat Rep*. 1980; 64: 519-524.
41. Thall PF, Nguyen HQ, Braun TM and Qazilbash MH. Using joint utilities of the times to response and toxicity to adaptively optimize schedule-dose regimens. *Biometrics*. 2013; 69: 673-682.
42. Trippa L, Rosner GL and Müller P. Bayesian enrichment strategies for randomized discontinuation trials. *Biometrics*. 2012; 68: 203-211.
43. Zhao L, Taylor JM and Schuetze SM. Bayesian decision theoretic two-stage design in phase II clinical trials with survival endpoint. *Stat Med*. 2012; 31: 1804-1820.
44. Zhao L and Woodworth G. Bayesian decision sequential analysis with survival endpoint in phase II clinical trials. *Stat Med*. 2009; 28: 1339-1352.
45. Berry DA and Ho CH. One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics*. 1988; 44: 219-227.
46. Chen MH and Willan AR. Determining optimal sample sizes for multistage adaptive randomized clinical trials from an industry perspective using value of information methods. *Clinical Trials*. 2013; 10: 54-62.
47. Cheng Y and Berry DA. Optimal adaptive randomized designs for clinical trials. *Biometrika*. 2007; 94: 673-687.
48. Cheng Y and Shen Y. Bayesian adaptive designs for clinical trials. *Biometrika*. 2005; 92: 633-646.
49. Cheng Y and Shen Y. An efficient sequential design of clinical trials. *Journal of Statistical Planning and Inference*. 2013; 143: 283-295.
50. Cheng Y, Su F and Berry DA. Choosing sample size for a clinical trial using decision analysis. *Biometrika*. 2003; 90: 923-936.
51. Chernoff H and Petkau AJ. Sequential medical trials involving paired data. *Biometrika*. 1981; 68: 119-132.
52. Claxton K and Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ*. 1996; 5: 513-524.
53. Claxton K and Thompson KM. A dynamic programming approach to the efficient design of clinical trials. *J Health Econ*. 2001; 20: 797-822.
54. Eckermann S and Willan AR. Expected value of information and decision making in HTA. *Health Econ*. 2007; 16: 195-209.

55. Gittins J and Pezeshk H. How large should a clinical trial be? *Journal of the Royal Statistical Society: Series D (The Statistician)*. 2000; 49: 177-187.
56. Gittins J and Pezeshk H. A behavioral Bayes method for determining the size of a clinical trial. *Drug Inf J*. 2000; 34: 355-363.
57. Gittins JC and Pezeshk H. A decision theoretic approach to sample size determination in clinical trials. *J Biopharm Stat*. 2002; 12: 535-551.
58. Halpern J, Brown Jr BW and Hornberger J. The sample size for a clinical trial: A Bayesian-decision theoretic approach. *Stat Med*. 2001; 20: 841-858.
59. Heitjan DF, Houts PS and Harvey HA. A decision-theoretic evaluation of early stopping rules. *Stat Med*. 1992; 11: 673-683.
60. Hornberger J and Eghtesady P. The cost-benefit of a randomized trial to a health care organization. *Control Clin Trials*. 1998; 19: 198-211.
61. Jennison C and Turnbull BW. Interim monitoring of clinical trials: Decision theory, dynamic programming and optimal stopping. *Kuwait Journal of Science*. 2013; 40: 43-59.
62. Jiang F, Jack Lee J and Müller P. A Bayesian decision-theoretic sequential response-adaptive randomization design. *Stat Med*. 2013; 32: 1975-1994.
63. Kikuchi T and Gittins J. A behavioral Bayes method to determine the sample size of a clinical trial considering efficacy and safety. *Stat Med*. 2009; 28: 2293-2306.
64. Kikuchi T and Gittins J. A behavioural Bayes approach to the determination of sample size for clinical trials considering efficacy and safety: Imbalanced sample size in treatment groups. *Stat Methods Med Res*. 2011; 20: 389-400.
65. Kikuchi T, Pezeshk H and Gittins J. A Bayesian cost-benefit approach to the determination of sample size in clinical trials. *Stat Med*. 2008; 27: 68-82.
66. Lewis RJ and Berry DA. Group sequential clinical trials: A classical evaluation of Bayesian decision-theoretic designs. *Journal of the American Statistical Association*. 1994; 89: 1528-1534.
67. Lewis RJ, Lipsky AM and Berry DA. Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: A frequentist evaluation. *Clinical Trials*. 2007; 4: 5-14.
68. Maroufy V, Marriott P and Pezeshk H. An optimization approach to calculating sample sizes with binary responses. *J Biopharm Stat*. 2014; 24: 715-731.
69. Mehta CR and Patel NR. Adaptive, group sequential and decision theoretic approaches to sample size determination. *Stat Med*. 2006; 25: 3250-3269.
70. Orawo LA and Christen JA. Bayesian sequential analysis for multiple-arm clinical trials. *Statistics and Computing*. 2009; 19: 99-109.
71. Patel NR and Ankolekar S. A Bayesian approach for incorporating economic factors in sample size design for clinical trials of individual drugs and portfolios of drugs. *Stat Med*. 2007; 26: 4976-4988.
72. Patel NR, Ankolekar S, Antonijevic Z and Rajcic N. A mathematical model for maximizing the value of phase 3 drug development portfolios incorporating budget constraints and risk. *Stat Med*. 2013; 32: 1763-1777.

73. Pezeshk H and Gittins J. A fully Bayesian approach to calculating sample sizes for clinical trials with binary responses. *Drug Inf J.* 2002; 36: 143-150.
74. Pezeshk H, Nematollahi N, Maroufy V and Gittins J. The choice of sample size: A mixed Bayesian/frequentist approach. *Stat Methods Med Res.* 2009; 18: 183-194.
75. Pezeshk H, Nematollahi N, Maroufy V, Marriott P and Gittins J. Bayesian sample size calculation for estimation of the difference between two binomial proportions. *Stat Methods Med Res.* 2013; 22: 598-611.
76. Wang YG. Gittins indices and constrained allocation in clinical trials. *Biometrika.* 1991; 78: 101-111.
77. Wathen JK and Christen JA. Implementation of backward induction for sequentially adaptive clinical trials. *Journal of Computational and Graphical Statistics.* 2006; 15: 398-413.
78. Wathen JK and Thall PF. Bayesian adaptive model selection for optimizing group sequential clinical trials. *Stat Med.* 2008; 27: 5586-5604.
79. Willan AR. Optimal sample size determinations from an industry perspective based on the expected value of information. *Clinical Trials.* 2008; 5: 587-594.
80. Willan AR and Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ.* 2010; 19: 549-561.
81. Willan AR and Eckermann S. Value of information and pricing new healthcare interventions. *Pharmacoeconomics.* 2012; 30: 447-459.
82. Willan AR and Kowgier M. Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clinical Trials.* 2008; 5: 289-300.
83. Willan AR and Pinto EM. The value of information and optimal clinical trial design. *Stat Med.* 2005; 24: 1791-1806.
84. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials.* 1989; 10: 1-10.
85. O'Hagan A and Stevens JW. Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Med Decis Making.* 2001; 21: 219-230.
86. Graf AC, Posch M and Koenig F. Adaptive designs for subpopulation analysis optimizing utility functions. *Biometrical Journal.* 2015; 57: 76-89.
87. Simon N and Simon R. Adaptive enrichment designs for clinical trials. *Biostatistics.* 2013; 14: 613-625.
88. Wang S-J, James Hung HM and O'Neill RT. Adaptive patient enrichment designs in therapeutic trials. *Biometrical Journal.* 2009; 51: 358-374.
89. Cheung YK. *Dose Finding by the Continual Reassessment Method.* Boca Raton, Fla. : Chapman and Hall, 2011.
90. Simes RJ. Application of statistical decision theory to treatment choices: implications for the design and analysis of clinical trials. *Stat Med.* 1986; 5: 411-420.

91. Grieve AP and Krams M. ASTIN: A Bayesian adaptive dose-response trial in acute stroke. *Clinical Trials*. 2005; 2: 340-351.
92. Berry DA and Fristedt B. *Bandit Problems: Sequential Allocation of Experiments*. London: Chapman and Hall, 1985.
93. Gittins J. *Multi-armed Bandit Allocation Indices*. Chichester: John Wiley, 1989.
94. Gittins JC, Weber R and Glazebrook KD. *Multi-armed Bandit Allocation Indices*. Hoboken, NJ: John Wiley & Sons, 2011.
95. Chaloner K, Church T, Louis TA and Matts JP. Graphical elicitation of a prior distribution for a clinical trial. *Journal of the Royal Statistical Society Series D (The Statistician)*. 1993; 42: 341-353.
96. Kadane J and Wolfson LJ. Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1998; 47: 3-19.
97. O'Hagan A. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1998; 47: 21-35.
98. Blanck TJJ, Conahan TJ, Merin RG, Prager RL and Richter JJ. Being an expert. In: Kadane JB, (ed.). *Bayesian Methods and Ethics in a Clinical Trial Design*. Wiley, 1996, p. 159-162.
99. Kinnersley N and Day S. Structured approach to the elicitation of expert beliefs for a Bayesian-designed clinical trial: a case study. *Pharm Stat*. 2013; 12: 104-113.
100. Parmigiani G and Inoue L. *Decision Theory: Principles and Approaches*. Wiley, 2009.
101. Lindley DV. *Making Decisions*. Wiley, 1971.
102. Emrich LJ and Sedransk N. Whether to participate in a clinical trial: The patient's view. In: Kadane JB, (ed.). *Bayesian Methods and Ethics in a Clinical Trial Design*. Wiley, 1996, p. 267-305.

Appendix A.1. Terms used in Scopus search with number of hits.^a

Search terms	Number of hits
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("decision theor*")) AND DOCTYPE(ar OR cp)	262
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY(sequential)) AND DOCTYPE(ar) AND (PUBYEAR > 2004)	1189
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY(sequential)) AND DOCTYPE(ar) AND (PUBYEAR < 2005)	1010
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY(sequential)) AND DOCTYPE(cp)	336
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("decision mak*")) AND DOCTYPE(ar) AND (PUBYEAR > 2008)	1705
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("decision mak*")) AND DOCTYPE(ar) AND (LIMIT-TO(PUBYEAR,2008) OR LIMIT-TO(PUBYEAR,2007) OR LIMIT-TO(PUBYEAR,2006) OR LIMIT-TO(PUBYEAR,2005))	901
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("decision mak*")) AND DOCTYPE(ar) AND (PUBYEAR < 2005)	1446
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("decision mak*")) AND DOCTYPE(cp)	1130
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY(bayes* PRE/5 decision)) AND DOCTYPE(ar OR cp)	70
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("optimal* design")) AND DOCTYPE(ar OR cp)	222
(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("optimal* decision")) AND DOCTYPE(ar OR cp)	16

(TITLE-ABS-KEY("small trial") OR TITLE-ABS-KEY("phase II") OR TITLE-ABS-KEY(pilot) OR TITLE-ABS-KEY("serie* of trial") AND TITLE-ABS-KEY("optimal* sample siz*")) AND DOCTYPE(ar OR cp)	10
TITLE-ABS-KEY("trial design") AND TITLE-ABS-KEY("decision theor*") AND DOCTYPE(ar OR cp)	25
TITLE-ABS-KEY("trial design") AND TITLE-ABS-KEY(bayes* PRE/5 decision) AND DOCTYPE(ar OR cp)	32
TITLE-ABS-KEY("trial design") AND TITLE-ABS-KEY("optimal* design") AND DOCTYPE(ar or cp)	38
TITLE-ABS-KEY("trial design") AND TITLE-ABS-KEY("optimal* decision") AND DOCTYPE(ar or cp)	4
TITLE-ABS-KEY("trial design") AND TITLE-ABS-KEY("optimal* sample size") AND DOCTYPE(ar or cp)	9
TITLE-ABS-KEY("trial design") AND TITLE-ABS-KEY(sequential* PRE/5 decision*) AND DOCTYPE(ar or cp)	3
(TITLE-ABS-KEY("clinical trial*") AND TITLE-ABS-KEY("decision theor*")) AND DOCTYPE(ar or cp)	278
(TITLE-ABS-KEY("clinical trial*") AND TITLE-ABS-KEY("optim* sample size*")) AND DOCTYPE(ar or cp)	65

^a Some search terms led to more than 2000 hits and to circumvent Scopus citation export restriction, these terms were split by publication year.