DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
2015, Vol. 49(1) 155-162

DOI: 10.1177/2168479014533970
tirs.sagepub.com

Bayesian Design of Proof-of-Concept Trials

Roland Fisch, PhD[1], Ieuan Jones, BSc[1], Julie Jones, MSc[1], Jouni Kerman, PhD[1], Gerd Karl Rosenkranz, PhD[1], and Heinz Schmidli, PhD[1]

Abstract

The proof-of-concept (PoC) decision is a key milestone in the clinical development of an experimental treatment. A decision is taken on whether the experimental treatment is further developed (GO), whether its development is stopped (NO-GO), or whether further information is needed to make a decision. The PoC decision is typically based on a PoC clinical trial in patients comparing the experimental treatment with a control treatment. It is important that the PoC trial be designed such that a GO/NO-GO decision can be made. The present work develops a generic, Bayesian framework for defining quantitative PoC criteria, against which the PoC trial results can be assessed. It is argued that PoC criteria based solely on significance testing versus the control are not appropriate in this decision context. A dual PoC criterion is proposed that includes assessment of superiority over the control and relevance of the effect size and hence better matches clinical decision making. The approach is illustrated for 2 PoC trials in cystic fibrosis and psoriasis.

Keywords

Bayesian inference, clinical trials, decision criteria, operating characteristics, phase IIa, proof of concept

Introduction

Clinical drug development typically proceeds in 3 phases, up to the submission for authorization by health authorities.[1] Phase I identifies the highest dose that is considered safe and tolerable in humans. Phase II provides initial evidence of efficacy in patients (phase IIa) as well as information on the dose-response relationship (phase IIb). Phase III trials are large, confirmatory trials conducted to deliver proof of safety and efficacy for regulatory submission, according to the requirements of health authorities.

In the first phase II trial (phase IIa), initial readings of efficacy in patients are obtained for the first time: That is, the concept of drug action is first tested in a patient population. For this reason, such trials are often called proof-of-concept (PoC) trials. The PoC decision point, right after the PoC trial, can be seen as the key milestone that separates exploratory from full development. The Pharmaceutical Research and Manufacturers of America (PhRMA) position paper[2] characterizes PoC as "the earliest point in the drug development process at which the weight of evidence suggests that it is 'reasonably likely' that the key attributes for success are present and the key causes of failure are absent" (based on a quote by Prof Paul Rolan, University of Adelaide).

Consequences of taking false PoC decisions are severe: A false-positive PoC decision (ie, declaring positive PoC whereas in truth the drug does not work as intended) means investment of resources in developing an inferior drug and undesirable exposure of patients to such a drug. Although drugs with inferior efficacy will often fail subsequently, allowing too many of them to pass a PoC decision point can be cumulatively extremely expensive.[3] False-negative PoC decisions (ie, declaring negative PoC whereas in truth the drug works as intended) affect both the sponsor (missed business opportunity) and future patients who could have benefitted from the drug. The consequences of false-positive and false-negative PoC decisions will likely differ among compounds and disease areas. These choices are portfolio (risk) decisions that should be agreed upon between teams and the appropriate governance groups, taking the competitive landscape into account.[4]

A PoC trial (ie, a trial to support a PoC decision) should be designed to allow informed decision making. To achieve this, quantitative PoC criteria are defined that should reflect the

[1] Development Biometrics & Statistical Science, Novartis Pharma AG, Basel, Switzerland

Submitted 22-Jan-2014; accepted 09-Apr-2014

Corresponding Author:
Roland Fisch, Novartis Pharma AG, Postfach, 4002 Basel, Switzerland.
Email: roland.fisch@novartis.com

Table 1. Four possible outcomes of the dual proof-of-concept (PoC) criterion.

	Evidence of Relevance (Condition 2 Holds)	Not Enough Evidence of Relevance (Condition 2 Does Not Hold)
Evidence of significance (condition 1 holds)	GO: evidence of sufficiently high effect	INDETERMINATE: evidence of marginal effect
Not enough evidence of significance (condition 1 does not hold)	INDETERMINATE: data too variable to make a clear decision	NO-GO: evidence of poor effect

main components of the PoC decision to be made, that is, considering both observed effect size and evidence for superiority against the control.[5-7] The PhRMA position paper[2] gives an overview of points to consider in this respect. Such PoC criteria can never fully capture the body of evidence leading to an actual PoC decision. However, the operating characteristics of the PoC criteria can be evaluated for various clinical trial designs until an acceptable design is found, given the available resources.

The next 2 sections of this paper define formal, quantitative PoC criteria, which should ideally reflect the main components of the PoC decision to be made. In the following, 2 typical examples of PoC trials are introduced. Some of the trial design options and optimization, as well as the use of historical control data, are discussed, based on the evaluation of operating characteristics, and these concepts are applied to the examples. Key features of proposed PoC criteria, trial design, and operating characteristics are summarized in the final section. Moreover, generalizations of the concept are referred to, such as adaptive trial designs, dose-response modeling, and the use of historical control data.

Formulating the Proof-of-Concept Criterion

A PoC trial, by definition, is a trial to support a PoC decision. Its primary objective is to evaluate a PoC criterion, which should reflect the main components of the PoC decision to be made.

In its simplest form, the PoC criterion will be based on one efficacy assessment, as the key attribute for success, where the parameter of interest is an improvement in efficacy, relative to placebo. However, statistical significance versus placebo alone will in general not be an appropriate PoC criterion. In most indications, drugs with minor average improvement relative to placebo will not be competitive even if found to be statistically significant. For example, consider cystic fibrosis, where lung function (ie, the forced expiratory volume in 1 second [FEV_1]) is a primary outcome. A minor observed improvement relative to placebo (eg, by 2%) may not be considered worthwhile, even if proven to be statistically significant.

The assumption in this paper is that a threshold can be agreed on below which the drug is not worth promoting into full development. We name this threshold separating marginal from competitive efficacy the "target difference" (TD); it may also be called "minimum clinically important difference" (as in Chuang-Stein et al[6]), "critical threshold" (as in Neuenschwander et al[5]), "required difference" (as in Cartwright et al[2]), "boundary of indifference," or "walk away point." Note that TD is in general smaller than both the treatment effect of the best compounds currently on the market and the "alternative" used in traditional power calculations. That is, TD will be determined from a multitude of factors—clinical, regulatory, and commercial—and this decision may be supported by evaluations of operating characteristics, such as the ones outlined below. In the FEV_1 example, TD may be an observed 5% improvement.

Taking the threshold TD into account, the PoC criterion becomes dual, where one component constitutes superiority versus placebo ("Does the drug work at all?"), and the other reflects the effect size relative to TD ("Does the drug have a relevant clinical effect?"). In other words, the PoC criterion should hold if and only if both of the following hold:

1. Significance: high confidence that the effect of investigational drug relative to placebo is bigger than zero
2. Relevance: moderate confidence that the drug effect, relative to placebo, is larger than the TD

These 2 conditions, when evaluated, give rise to 4 possible outcomes, as depicted in Table 1. The 2 fields marked INDETERMINATE in Table 1 indicate that one component of the PoC criterion holds, the other not.

Further action will depend on the specifics of the project. Often, if there is evidence for a relevant effect but not enough to make the significance threshold, a trial expansion may be considered. Evidence for significance but not relevance may suggest reconsidering the target population. For example, the observed effect may be relevant in a subpopulation, and the decision may be to redirect further development to this subpopulation.

Quantitative Bayesian PoC Criterion

In many instances, the dual PoC criterion can be formulated within the classical hypothesis testing framework. However,

we choose here a Bayesian formulation, due to its greater flexibility.

In a Bayesian framework, the uncertainty about parameters indexing the distribution of the data is expressed via probability distributions. Bayes' theorem allows derivation of these probability distributions, given the likelihood function (ie, the statistical model) and a prior distribution for the parameters. Bayesian statistical methods have become an indispensable tool in clinical research and development.[8-11] With the advent of Markov chain Monte Carlo approaches,[12] Bayesian analyses have become feasible for a wide set of statistical models using standard statistical software.

Assume that the parameter θ is a measure of effect. In an indication where lung function is a primary outcome (eg, measured by FEV_1), θ could be the difference between treated and placebo group, in % FEV_1 improvement. In the Bayesian paradigm, a posterior probability distribution of θ given the data, $p(\theta \mid \text{data})$, can be obtained, given the likelihood $l(\theta \mid \text{data})$, and a prior distribution $p(\theta)$:

$$p(\theta \mid \text{data}) \propto 1(\theta \mid \text{data}) \times p(\theta) \,(\text{Bayes}' \text{theorem}).$$

The prior distribution $p(\theta)$ may be chosen to be "uninformative" if no relevant prior information exists. If, however, historical information is available (eg, on the effect in the control group), the Bayesian approach allows incorporation of this as an "informative" prior. The method to derive informative priors using a meta-analytic-predictive approach is described by Neuenschwander et al[13] and Gsteiger et al[14]; the psoriasis example given below demonstrates the potential benefits.

In the previous section, the PoC criterion was verbally defined as a dual one, consisting of a "significance" and a "relevance" component. Using the Bayesian formulation, this dual PoC criterion translates to the following:

1. Significance: $\Pr(\theta > 0 \mid \text{data}) \geq 1 - \alpha$
2. Relevance: $\Pr(\theta > \text{TD} \mid \text{data}) \geq \gamma$,

where Pr stands for probability; $1 - \alpha$ stands for the high level of confidence we desire in the significance component (eg, $1 - \alpha = 0.9$); and γ stands for the moderate level of confidence in a relevant effect, bigger than TD (eg, $\gamma = 0.50$).

Figure 1 illustrates the use of these criteria graphically: The PoC criterion is fulfilled if and only if the shaded area is bigger than $1 - \alpha = 0.9$ and the solid area exceeds $\gamma = 0.5$.

PoC Trial Examples

Two examples are introduced, one discussing a PoC trial in cystic fibrosis with a continuous endpoint (FEV_1), one on a PoC trial in psoriasis with a binary endpoint (PASI75). In both examples, we assume that first-in-human, healthy volunteer, single-dose and multiple-dose trials have been carried out, such that a safe dose for a first trial in patients could be determined. The next trial to be planned is the proof-of-concept trial, to support an informed PoC decision. Whereas both examples are derived from actual clinical development projects, they are both somewhat simplified to illustrate more clearly the main ideas.

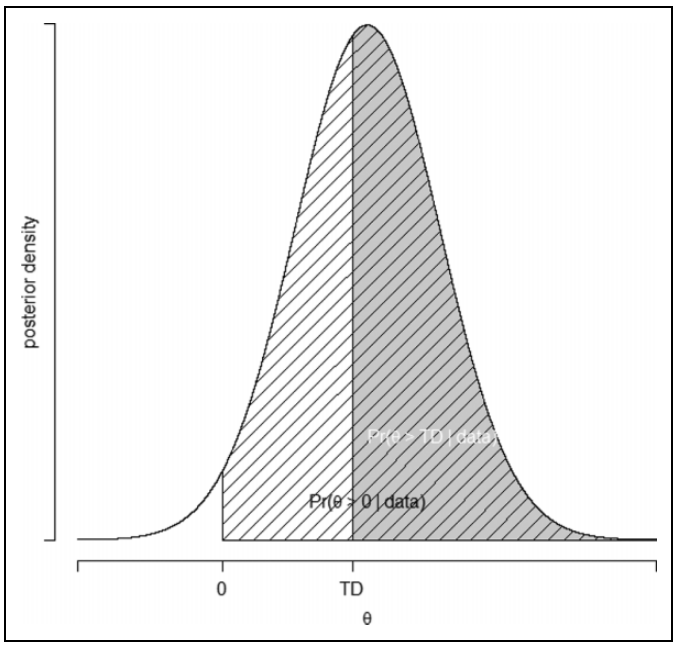

Figure 1. Illustration of the dual proof-of-concept (PoC) criterion.

PoC in Cystic Fibrosis

Consider a clinical development project in cystic fibrosis, at the stage of planning a PoC trial. The planned design is a randomized, double-blind, placebo-controlled, parallel group trial. The primary endpoint of interest is the change from baseline in percentage of predicted FEV_1 at day 28. We first consider whether the drug effect over placebo is bigger than zero: If there is a 90% chance in this study that the true drug effect is indeed better than placebo, then the team would consider this appropriate evidence that the effect is statistically of interest.

However, historical data from a competitor have already shown mean improvements over placebo of around 10% in this endpoint in a similar population. After discussions with the team it was believed that if the magnitude of the true effect of the drug was below 5%, then this would definitely not be good enough to progress drug development for this disease. Thus, the TD was set at 5% improvement in FEV_1, relative to placebo. The level of confidence to exceed the TD was set to 0.5. Therefore, with θ the difference between treated and placebo group, in % FEV_1 improvement, the quantitative PoC dual criterion is as follows:

1. Significance: $\Pr(\theta > 0 \mid \text{data}) \geq 0.9$.
2. Relevance: $\Pr(\theta > 5\% \mid \text{data}) \geq 0.5$.

PoC in Psoriasis

In this study, a new treatment for psoriasis is being assessed. The planned design is a randomized, placebo-controlled, double-blind, parallel group study. In this case the primary efficacy endpoint is proportion of patients with at least a 75% reduction from baseline in the PASI (Psoriasis Area and Severity Index) at 12 weeks, the PASI75 rate.

The existing treatment already has a meaningful benefit over placebo, so for the further development of the new treatment, the team requires evidence that the new treatment works at least as well as the existing treatment. There is no desire to invest in the development of a treatment that does not offer a meaningful benefit to patients over the existing treatments. Analysis of data from the alternative treatment shows that this treatment has a difference compared with placebo of around 40 percentage points in the PASI75 rates. Thus, the chosen PoC criteria were as follows:

1. Significance: $\Pr(\theta > 0 \mid \text{data}) \geq 0.9$
2. Relevance: $\Pr(\theta > 0.40 \mid \text{data}) \geq 0.5$,

where θ is the difference in the PASI75 rate between the new treatment and placebo.

Design Optimization and Operating Characteristics

A good PoC trial design will obviously be one that leads to few false decisions (false positive or false negative) and keeps cost and duration of the trial at a reasonable level. By an appropriate choice of design, one can achieve an acceptable relation between the probability of correct GO or NO-GO decisions and the probability of an INDETERMINATE outcome. Design optimization will consist of evaluating a range of design options with respect to "operating characteristics," that is, the chances of making right or wrong decisions, given assumptions of true effect. The operating characteristics of a specific design are usually evaluated by clinical trial simulations. Assuming different sizes of treatment effects, data for a large number of clinical trials are simulated, and for each of these virtual trials, the PoC criteria are evaluated. The percentage of simulated trials for which the PoC outcome is GO, NO-GO, or INDETERMINATE is then a measure to compare the properties of different trial designs.

PoC in Cystic Fibrosis

In this cystic fibrosis PoC trial, a parallel group design was chosen; for reasons unrelated to the PoC criteria, a 2:1 active-placebo randomization rate was used. Based on evidence from previous trials, the between-subject standard deviation is assumed to be 10 (on the % FEV_1 scale). Initially, a total sample size of 78 was considered (ie, 52 patients on active treatment and 26 on placebo). Note that if a hypothesis testing approach was used, testing against the null hypothesis of no improvement over placebo, such a trial would have 80% power to detect a difference of 5% in predicted FEV_1, using a 1-sided test at the 10% level. Thus, if the true drug effect was equal to placebo, then 10% of trials would erroneously declare statistical significance over placebo; if the true effect in the active treatment was 5% better than placebo, 80% of trials would correctly declare active superior.

In the following, the operating characteristics of the proposed dual PoC criterion (labeled "C" below) are compared with (A) an approach using a single criterion of statistical significance only and (B) a single criterion approach that examines the clinical relevance criterion only.

(A) Statistical Significance Criterion Only

First, consider a PoC criterion based on significance alone:

$$\Pr(\theta > 0\% \mid \text{data}) \geq 0.9,$$

where θ is the difference between treated and placebo group in % FEV_1 improvement. This is akin to a 1-sided hypothesis test at the 10% level.

Figure 2A shows operating characteristics (ie, the probabilities of declaring GO and NO-GO) depending on the true effect size. For a true drug effect of 4%, an erroneous GO decision would be declared in about 60% of trials (false-positive decision on marginal efficacy, ie, below 5%).

This approach therefore leads to too high a chance to declare positive PoC if efficacy is clearly below the TD. Decreasing the sample size or increasing the confidence level (eg, to 0.95) may improve on this (ie, reduce the chance of a GO decision with marginal efficacy); however, this will then also lead to a smaller chance of true positives (ie, of declaring GO correctly) in the presence of relevant efficacy.

(B) Relevance Criterion Only

Now, consider the operating characteristics of a criterion with just a relevance component:

$$\Pr(\theta > 5\% \mid \text{data}) \geq 0.5.$$

Figure 2B shows its operating characteristics, that is, the probabilities of declaring GO and NO-GO, depending on the true effect size. For a true drug effect of 4% (marginal efficacy), an erroneous GO decision would be declared in about 34% of trials. This is better than the significance-only criterion. Conversely, for a relevant drug effect of 6%, the probability of an erroneous NO-GO decision would be about 66%, which is much worse than the significance-only rule. The effect size

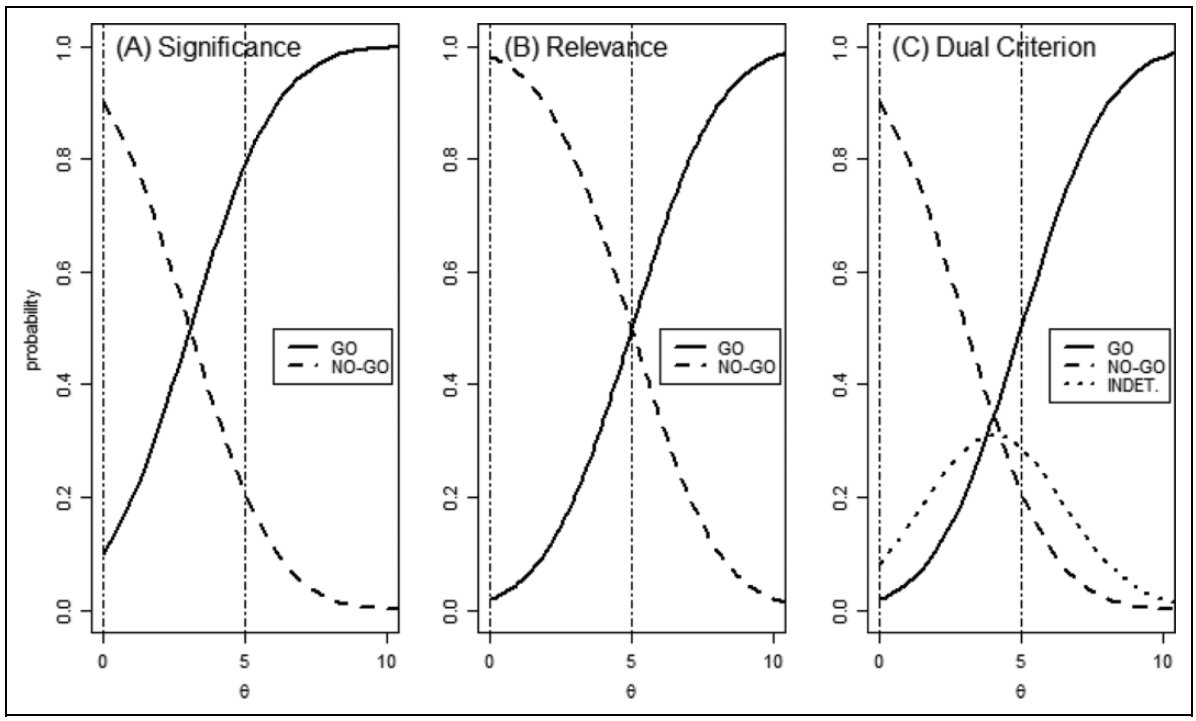

Figure 2. Cystic fibrosis example: operating characteristics of simple and dual proof-of-concept (PoC) criteria.

would have to be around 7% for an 80% chance of a GO decision. With increasing sample size, the operating characteristics will improve (ie, the chances of a GO decision will increase) if the true effect is beyond the TD of 5%, and the chances of a NO-GO decision will increase if the true effect is below the TD of 5%. However, through ignoring the "significance" component, we have no control over the significance level when comparing with placebo; if small sample sizes are chosen such that the standard error of the effect estimate is similar or bigger than the TD, a relatively high chance of a GO decision can result when treatment is not even superior to placebo.

(C) Dual Criterion

Consider now the dual quantitative PoC criterion:

1. Significance: $\Pr(\theta > 0 \mid \text{data}) \geq 0.9$.
2. Relevance: $\Pr(\theta > 5\% \mid \text{data}) \geq 0.5$.

Figure 2C shows the operating characteristics of the dual decision criteria mapped to 3 decision spaces, GO, NO-GO, and INDETERMINATE. The solid line shows the typical "power" curve for a GO decision, the dashed line represents a NO-GO, and the dotted line represents the indeterminate case. It is seen that the chance of erroneously declaring GO for a placebo-like drug is nearly zero, similar to the relevance-only rule but less than the significance-only rule that shows 10% when $\theta = 0$. For a true drug effect of only 4%, an erroneous GO decision would be declared in about 33% of trials. In this case, this is similar to the relevance-only criterion. With an effect of 6%, the probability of a erroneous NO-GO decision (dashed line) is 10%, which seems acceptably low; the probability of a GO is 68%. There is now a 22% chance of an INDETERMINATE decision because a true (unobserved) effect of 6% is close to the point of indifference of 5%. The effect size needs to be 7% for 80% chance of a GO.

Assuming the variance to be known, one can choose sample sizes for which the "significance" and the "relevance" component are equivalent (ie, the one occurs if and only if the other occurs) and no INDETERMINATE outcome can result. For smaller sample sizes, the "significance" component will dominate, and INDETERMINATE outcomes are most likely for true effect sizes just above TD. For large sample sizes, INDETERMINATE outcomes are most likely for true effect sizes between 0 and TD. This latter case is in fact chosen in this example, since the team may want to consider alternative ways to continue the clinical program, if the outcome is significant but below TD. For example, one may consider subpopulations of the cystic fibrosis indication, which is quite heterogeneous to start with. The task of the trial design team is to strike a reasonable balance between the chances of making false-positive, false-negative, and indeterminate decisions and the cost and duration of the trial; in this cystic fibrosis example, the chosen sample size may be considered feasible and the operating characteristics acceptable.

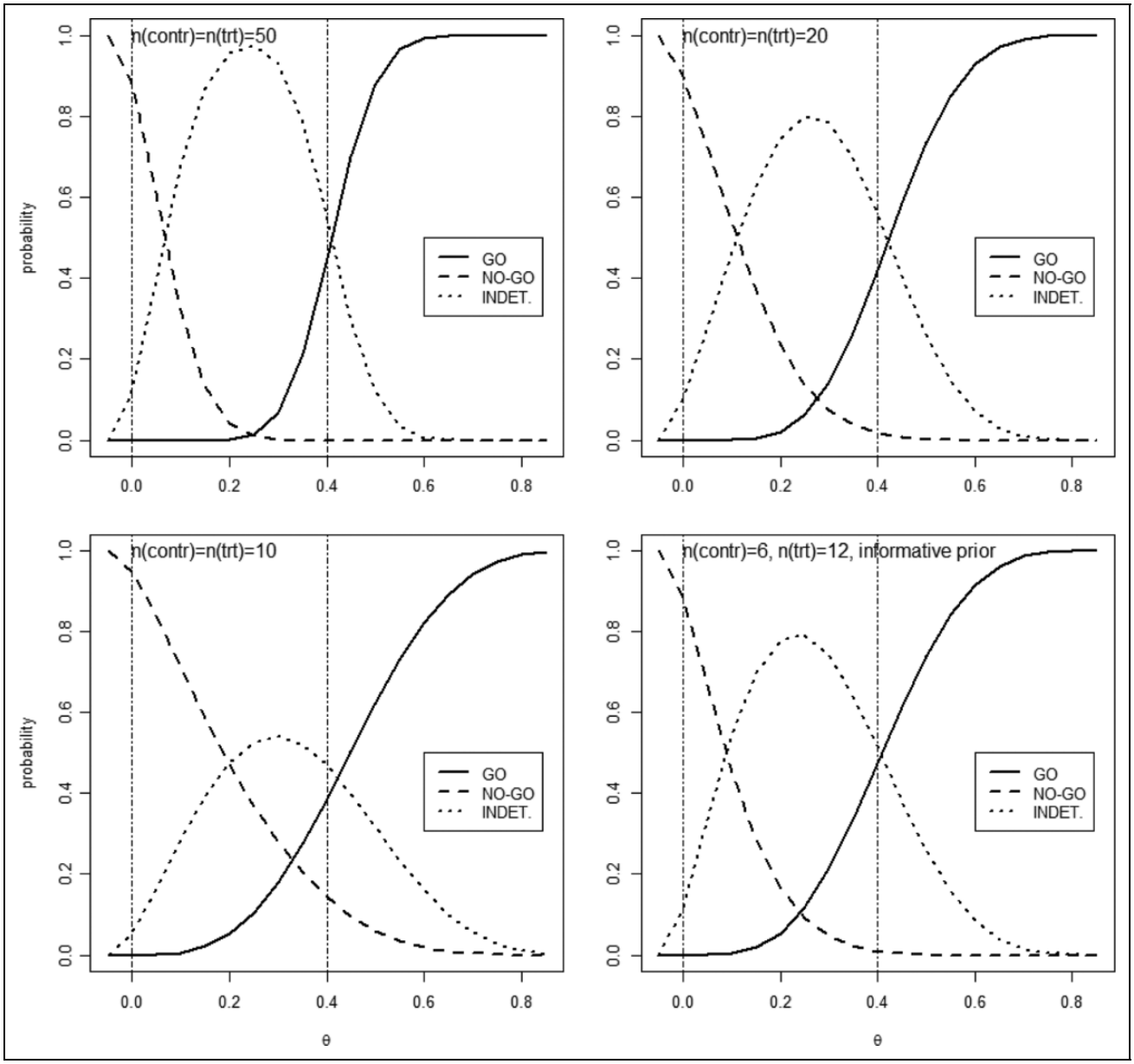

Figure 3. Psoriasis example: operating characteristics for different sample sizes and using an informative prior.

PoC in Psoriasis

The planned design in the psoriasis indication is a randomized, placebo-controlled, double-blind, parallel group study. The primary efficacy endpoint is the PASI75 rate, that is, the proportion of patients with at least a 75% reduction from baseline in the Psoriasis Area and Severity Index, at 12 weeks. The PoC criterion is as follows:

1. Significance: $\Pr(\theta > 0 \mid \text{data}) \geq 0.9$
2. Relevance: $\Pr(\theta > 0.40 \mid \text{data}) \geq 0.5$,

where θ is the difference in the PASI75 rate between the new treatment and placebo.

If we first consider equal sample sizes per group, with no interim analyses and no informative priors, the design space is defined by a single parameter—the sample size per group. The first 3 panels in Figure 3 show the probability of achieving the 3 different outcomes (GO, INDETERMINATE, NO-GO) depending on the true difference in response rate between active treatment and placebo, for sample sizes 50, 20, and 10 per group. The response rate for placebo is assumed to be 5% in this simulation.

Since the TD is large, the overall size of the study may be relatively small and sample sizes of no more than 50 patients per group are considered. These sizes lead to reasonable chances of coming to the correct conclusion. For example, if

the true response rate for the active treatment is 55% (ie, 50 percentage points away from placebo), then a sample size of 50 patients per group provides almost 90% probability of declaring GO. For this reason smaller sample sizes can be considered as shown in Figure 3 (ie, 20 or 10 per group).

When comparing the dotted lines in the 3 panels of Figure 3 for sample sizes 50, 20, and 10, we see that a larger sample size leads to a higher probability of being in the INDETERMINATE zone when the true treatment effect is less than the TD. However, the balance between the benefit of a larger sample size and its cost is not the same as in a standard power calculation. A lower sample size leads to an acceptable decrease in the probability of an INDETERMINATE outcome if the effect is positive but smaller than TD while still providing a satisfactory chance of a GO outcome if the effect is larger than TD.

The dashed lines show the probability of a NO-GO decision. It is important to consider what happens in the case of a game-changing treatment (a treatment that is much better than the TD). In this example, for a true treatment effect of 75% (70 percentage points away from placebo), a sample size of 50 patients per group leads to an approximately zero chance of declaring NO-GO. Lower sample sizes also show this same property. For true effect sizes closer to the target effect, however, the larger sample sizes perform better. If the true response rate for the active treatment is 65%, then a sample size of 50 patients per group provides less than 1% probability of declaring NO-GO whereas a sample size of 10 patients per group has a probability of 2%, which may be considered an acceptable false-negative rate for the assumed effect.

One may consider using an informative prior, since many historical trials in the same population with a similar design as the planned PoC trial are available, involving different experimental treatments but the same control (placebo). In a recent systematic review, Bansback and colleagues[15] identified 21 placebo-controlled studies in patients with moderate to severe plaque psoriasis, with a total of 3071 placebo patients. An informative prior for the placebo effect in the new PoC trial was derived from the historical data; it can be approximated by a beta-distribution, with a mean of 4.4%, and is worth about 103 patients. This is a considerable down-weighting of the historical placebo information due to between-trial heterogeneity. The bottom right panel in Figure 3 shows the operating characteristics of the design when using the informative prior for the placebo, allocating 12 patients to active treatment and 6 to placebo. It has characteristics superior to the 10+10 patients design and is similar to the 20+20 patients design. This emphasizes the benefit of using an informative prior if relevant historical data exist.

Discussion and Conclusions

Proof-of-concept trials should provide key information to decide whether it is worth continuing drug development. In this paper, we propose a strategy for the rational design and analysis of PoC trials. As a first step, a quantitative PoC criterion is defined that supports the clinical decision making at the end of the trial. This criterion typically contains 2 components, related to the level of evidence of superiority against placebo and to the observed effect size of the primary efficacy endpoint, respectively. The proposed dual criterion allows for the possibility of an indeterminate outcome; it accounts for the fact that if only 1 component of the dual criterion holds, a reasonable decision based on the PoC trial outcome can be to further explore, rather than a GO/NO-GO decision.

In formulating a PoC criterion, we used a Bayesian framework; in our experience this facilitates communication within the clinical team, as this is then more clearly seen as different from the traditional approach focusing solely on statistical significance. It should be noted that the classical statistical framework allows formulation of very similar criteria[16] should the Bayesian approach be considered a hurdle for implementation.

It is understood that a simple quantitative PoC criterion, as proposed here, can never fully reflect the complexity of the actual PoC decision to be made. However, it can still be valuable in many respects. First, it allows quantification of the risk of a false decision, which is not possible for qualitative criteria. Possibly of equal importance is that quantitative criteria can be used to optimize the design of the PoC trial through the evaluation of operating characteristics. An optimal trial design with respect to a quantitative PoC criterion is likely advantageous for supporting the PoC decision, assuming the PoC criterion is a reasonable surrogate for the PoC decision. Operating characteristics of various designs—ie, "what-if" evaluations—help to assess how likely it is that good decisions are derived, such that that drugs with relevant effects are likely to pass the PoC criterion and drugs with marginal effects are not. Design features such as the number of patients allocated to drug and control can then be chosen such that the operating characteristics are satisfactory.

To best illustrate our strategy, we considered here the simplest PoC designs only, where the experimental treatment is compared with a control treatment and where no adaptations are considered. In many PoC trials, interim analyses are included that may allow stopping the trial early for success or futility. With adaptive designs, fewer patients are needed on average than with fixed designs, as a trial can be stopped earlier if the treatment effect is either very small or very large.[9,11,17] Again, clinical trial simulations can be used to evaluate the operating characteristics and to compare different adaptive design options with each other and with fixed designs. The corresponding evaluation of operating characteristics is more complex but still quite feasible.[18] The R package gsbDesign[19] can be freely downloaded (from cran.r-project.org/web/packages/gsbDesign/) to aid with this. Additional levels of complexity may be considered, for example, by modeling the dose-response relationship within an adaptive PoC trial

design[17] or by using interim decision criteria based on predictive probabilities.[20]

Often, historical trials in the same population with a similar design as the planned PoC trial are available that involve different experimental treatments but the same control (eg, placebo). The method to derive prior information using a meta-analytic-predictive approach is described by Neuenschwander et al,[13] Gsteiger et al,[14] and Bansback et al[15]; a recent PoC trial using historical controls is described by Baeten et al.[21] As shown in the psoriasis example, such historical control information allows a reduction of the number of patients, which reduces cost and trial duration, facilitates recruitment, and may be more ethical in some situations.

The strategy for design and analysis of PoC trials proposed here is increasingly used in pharmaceutical companies (see, eg, Hueber et al[22]). We believe that this approach leads to more appropriate designs for PoC trials and to better decisions when the trial results become available.

Declaration of Conflicting Interests

All authors are employed by Novartis Pharma AG, Basel, Switzerland.

Funding

All authors are employed by Novartis Pharma AG, Basel, Switzerland.

References

1. Sheiner LB. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther*. 1997;61(3):275-291.
2. Cartwright ME, Cohen S, Fleishaker JC, et al. Proof of concept: a PhRMA position paper with recommendations for best practice. *Clin Pharmacol Ther*. 2010;87(3):278-285.
3. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*. 2010;87(3):272-277.
4. Mallinckrodt C, Molenberghs G, Persinger C, Ruberg S, Sashegyi A, Lindborg S. A portfolio-based approach to optimize proof-of-concept clinical trials. *J Biopharm Stat*. 2012;22(3):596-607.
5. Neuenschwander B, Rouyrre N, Hollaender N, Zuber E, Branson M. A proof of concept phase II non-inferiority criterion. *Stat Med*. 2011;30(13):1618-1627.
6. Chuang-Stein C, Kirby S, Hirsch I, Atkinson G. The role of the minimum clinically important difference and its impact on designing a trial. *Biopharm Stat*. 2011;10(3):250-256.
7. Chuang-Stein C, Kirby S, French J, et al. A quantitative approach for making go/no-go decisions in drug development. *Drug Inf J*. 2011;45:187-202.
8. Spiegelhalter DJ, Freedman LS, Parmar MKB. Applying Bayesian ideas in drug development and clinical trials. *Stat Med*. 1993;12 (15-16):1501-1511.
9. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health Care Evaluation*. New York, NY: John Wiley & Sons; 2004.
10. Hobbs BP, Carlin BP. Practical Bayesian design and analysis for drug and device clinical trials. *J Biopharm Stat*. 2007;18:54-80.
11. Berry SM, Carlin BP, Lee JJ, Müller P. *Bayesian Adaptive Methods for Clinical Trials*. London, UK: Chapman & Hall/CRC; 2010.
12. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc*. 1990;85:398-409.
13. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clin Trials*. 2010;7(1):5-18.
14. Gsteiger S, Neuenschwander B, Mercier F, Schmidli H. Using historical control information for the design and analysis of clinical trials with overdispersed count data. *Stat Med*. 2013;32(21): 3609-3622.
15. Bansback N, Sizto S, Sun H, Feldman S, Willian MK, Anis A. Efficacy of systemic treatments for moderate to severe plaque psoriasis: systematic review and meta-analysis. *Dermatology*. 2009;219(3):209-218.
16. Kieser M, Hauschke D. Assessment of clinical relevance by considering point estimates and associated confidence intervals. *Biopharm Stat*. 2005;4(2):101-107.
17. Smith MK, Jones I, Morris MF, Grieve AP, Tan K. Implementation of a Bayesian adaptive design in a proof of concept study. *Biopharm Stat*. 2006;5:39-50.
18. Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H. A practical guide to Bayesian group sequential designs. *Biopharm Stat*. 2014;13(1):71-80.
19. Gerber F, Gsponer T. gsbDesign: an R Package for evaluating operating characteristics for a group sequential Bayesian design. submitted to *J Stat Softw.* 2014.
20. Dmitrienko A, Wang M-D. Bayesian predictive approach to interim monitoring in clinical trials. *Stat Med*. 2006;25: 2178-2195.
21. Baeten D, Baraliakos X, Braun J, et al. Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2013;382(9906):1705-1713.
22. Hueber W, Sands BE, Lewitzky S, et al; Secukinumab in Crohn's Disease Study Group. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut*. 2012;61(12):1693-700.