PERSPECTIVES

A GUIDE TO DRUG DISCOVERY — OPINION

The future of drug development: advancing clinical trial design

John Orloff, Frank Douglas, Jose Pinheiro, Susan Levinson, Michael Branson, Pravin Chaturvedi, Ene Ette, Paul Gallo, Gigi Hirsch, Cyrus Mehta, Nitin Patel, Sameer Sabir, Stacy Springs, Donald Stanski, Matthias R. Evers, Edd Fleming, Navjot Singh, Tony Tramontin and Howard Golub

Abstract | Declining pharmaceutical industry productivity is well recognized by drug developers, regulatory authorities and patient groups. A key part of the problem is that clinical studies are increasingly expensive, driven by the rising costs of conducting Phase II and III trials. It is therefore crucial to ensure that these phases of drug development are conducted more efficiently and cost-effectively, and that attrition rates are reduced. In this article, we argue that moving from the traditional clinical development approach based on sequential, distinct phases towards a more integrated view that uses adaptive design tools to increase flexibility and maximize the use of accumulated knowledge could have an important role in achieving these goals. Applications and examples of the use of these tools — such as Bayesian methodologies — in early- and late-stage drug development are discussed, as well as the advantages, challenges and barriers to their more widespread implementation.

Pharmaceutical innovation is increasingly risky, costly and at times inefficient, which has led to decreased industry productivity[1–3]. Estimates for the average cost of bringing a new drug to market range between $800 million and $2 billion, in which late-stage failures and the rising costs of Phase II and III trials represent key components[4–9]. Conducting these phases of development more effectively and reducing attrition rates are therefore major goals. The problem of attrition is particularly acute in Phase II trials[10], owing to factors such as the lack of proof of relevance for the biological target in a given disease intervention and insufficient understanding of the dose–response relationship of the new molecular entity.

As recognized by the US Food and Drug Administration (FDA) Critical Path Initiative, novel approaches to clinical trial and programme design could have a key role in overcoming these challenges. The traditional approach to drug development separates clinical development into sequential, distinct phases, in which progress is measured at discrete milestones, separated by 'white space'. We argue that the effectiveness of the clinical development can be improved by adopting a more integrated model that increases flexibility and maximizes the use of accumulated knowledge. In this model, broader, more flexible phases leading to submission for approval are designated 'exploratory' and 'confirmatory' (FIG. 1). This model is adaptive, parallel and data-led, and allows all available knowledge to be appropriately shared across the breadth of development studies to improve the quality, timeliness and efficiency of the process.

Central to this model of drug development are novel tools, including modelling and simulation, Bayesian methodologies, and adaptive designs, such as seamless adaptive designs and sample-size re-estimation methods (BOX 1). These can ensure the judicious use of limited patient resources, reduce patient exposure to ineffective or poorly tolerated doses, and lead to the recruitment of patients who, on the basis of biomarker analysis, are most likely to respond and those with the most favourable benefit/risk ratio.

In this article, we describe the general issues and methods involved, and illustrate how the tools can be applied in both exploratory and confirmatory drug development by using specific cases in which modern trial designs and statistical approaches have been successful. We hope to raise awareness of these issues among those involved in clinical trials and provide guidelines to ensure that the most appropriate solutions are implemented, with the ultimate goal of increasing the efficiency and probability of success in clinical development.

Exploratory phase of development

Modelling is a key feature of the more integrated approach to drug development (FIG. 1). Biological modelling is used to understand genetic, biochemical and physiological networks, as well as pathways and processes underlying disease and pharmacotherapy[11,12]. Pharmacological modelling guides clinical trial design, dose selection and development strategies[13,14]. Finally, statistical modelling can be used to assess development strategies and trial designs in populations[11,12,15]. These three types of modelling should be used throughout the drug development process to maximize their impact and synergies.

In the exploratory phase, modelling and simulation can help refine dose selection and study design. Early development studies are conducted with fairly restricted resources (duration, sample sizes and so on), and the use of all available information is crucial for effective decision making[16]. However, it should be noted that early development decisions based on biomarkers that have not been fully qualified can be misguided if such biomarkers eventually do not prove to correlate with, or be predictive of, the final outcome. Accordingly, it is important to conduct methodology research in parallel to the development programme to establish the correlation between the biomarker and late-stage endpoints or outcomes.

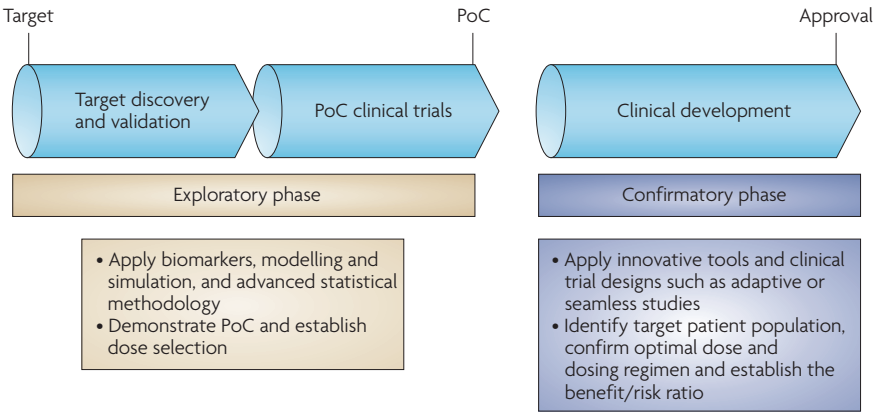


Figure 1 | **A novel model for clinical development.** During the exploratory phase of development, this model uses all available knowledge and tools, including biomarkers, modelling and simulation, as well as advanced statistical methodology. Trials are designed to determine proof-of-concept (PoC) and to establish dose selection to a level of rigour that will enhance the likelihood of success in the confirmatory phase. During the confirmatory phase, modern designs, tools and knowledge are applied to larger-scale studies with the goal of identifying the target patient population in which the drug is efficacious, establishing the benefit/risk ratio and confirming the optimal dose and dosing regimen. During this phase, innovative clinical trial designs such as adaptive or seamless studies compress timelines, improve dose and regimen selection, and reduce the number of patients assigned to non-viable dosing regimens.

Modelling and simulation approaches can be used to represent dose–response and time–response behaviour of safety and efficacy endpoints. Furthermore, these approaches can be combined with Bayesian methods to provide a continuous flow of information across different phases of development. For example, preclinical data can be used to construct models and to provide prior information on model parameters. Likewise, the results from a proof-of-concept (PoC) study can be used to form prior distributions for a similar model to be used in a subsequent dose-finding study[11,12,17,18].

An additional benefit of modelling in early development is that it allows the use of external information (for example, baseline values for safety endpoints) to estimate characteristics of interest about the population. Given the vast quantity of data from other development programmes that are available in most pharmaceutical companies, as well as current discussions within the industry about sharing placebo data across companies, this has huge potential for improving the efficiency of investigation in early development.

Modelling and simulation for dose and dose regimen selection. An important goal of a drug development programme is the selection of a dose and dosing regimen that achieves the target clinical benefit while minimizing undesirable adverse effects. Biological and pharmacological modelling can be very useful in this context[19,20]. For example, we (J.O., J.P., M.B., P.G. and D.S.) have used such modelling in the dose selection for canakinumab (Ilaris; Novartis), a monoclonal antibody that has recently been approved for the treatment of the rare genetic disease Muckle–Wells syndrome (FIG. 2). Clinical data on the relationship between activity of the therapeutic target (interleukin-1), markers of inflammation and remission of symptoms were captured in a mathematical model that was continuously adjusted to fit emerging data. Simulation was then used to propose a suitable dose and dosing regimen to achieve the desired response for the majority of patients — in this instance, an 80% probability that 90% of patients would remain flare-free for 2 months. The data derived from this modelling exercise allowed for selection of a dosing regimen that was investigated and confirmed in a Phase III trial[21] (clinical data on various dosing intervals provided the raw data for the modelling and simulation exercise that finalized the dose and regimen selection for Phase III). Similarly, modelling has been used to predict the impact of changing the dose or dosing regimen of a dipeptidyl peptidase IV inhibitor that is being developed for the treatment of type 2 diabetes (see Supplementary information S1 (box)).

Bayesian modelling combined with use of external baseline data to improve efficacy and safety signal detection in early development. Early development studies for establishing PoC often use small patient cohorts (10–20 subjects). These patients are typically observed for a relatively short period of time (several weeks) to evaluate early efficacy and safety signals, which are frequently measured on a continuous scale and observed several times over the duration of the study. However, the endpoints for the decision to proceed with development or not are typically based on a single time point (for example, change from baseline at the end of the study) and use dichotomized versions of the original variables to characterize responder and non-responder behaviour. An example of the latter is the transformation of continuous liver function test measurements (for example, alanine aminotransferase (ALT) and aspartate aminotransferase (AST)) into binary indicators (for instance, exceeding three times the upper limit of normal (ULN)). There are, therefore, two types of information loss that often occur in PoC studies: the dichotomization of continuous endpoints and a failure to use all of the available longitudinal measurements collected in the study[22].

A typical design for efficacy and safety evaluation in a PoC study is to use cohorts in a dose-escalation algorithm. Cohorts are assigned, in sequence, to increasing doses until the maximum tolerated dose is reached, or unacceptable safety is observed for a given cohort. A new cohort is only allowed to start once acceptable safety signals are verified for all previous doses. At the end of the study, the goal is to either determine a dose range for further exploration in Phase IIb, or to conclude that no PoC can be established based on the efficacy–safety trade-off.

Because of small cohort sizes, only safety problems occurring in a relatively large percentage of patients can be reliably detected by dose-escalation procedures. Likewise, only relatively strong efficacy signals can be detected with reasonable statistical power. The detection of safety and efficacy signals can be made more efficient in various ways: by drawing on data and information external to the trial, and by deploying longitudinal modelling approaches to make use of all available information. Furthermore, the utility of PoC studies within drug development programmes can be enhanced by incorporating the information obtained in them directly into later-phase trials[11,12]. Bayesian modelling techniques are particularly useful

Box 1 | **Tools, methods and designs for enhancing clinical development**

Here, we summarize some of the key tools, methods and designs that can be incorporated into the drug development process.

Modelling and simulation

These techniques are a cornerstone of the novel drug development model. In the exploratory phase, modelling and simulation can help refine dose selection and study design, and to represent dose–response and time–response behaviour of safety and efficacy endpoints. In combination with Bayesian methods, these can provide a continuous flow of information across different phases of development. Modelling in early development also enables the use of external information (an important issue in light of current discussions within the industry about sharing placebo data across companies), which could greatly increase the efficiency of investigations in early development.

In the confirmatory phase, simulation can clarify how different study designs affect the outcome and likelihood of success, thereby guiding development strategy. In the latter case, this is facilitated by pooling many sources of data both from prior studies of the drug and external data that might be an informative guide to achieve better decision-making. Furthermore, these techniques can be used not just during the trial-design process, but also mid-study through the use of adaptive trial designs.

Bayesian methodology

This relies on the use of probability models to describe knowledge about parameters of interest (for example, the treatment effect of a drug in development). Bayesian inference uses principles from the scientific method to combine prior beliefs with observed data, producing enhanced, updated information (for reviews, see REFS 22,23). Using Bayesian methodologies, initial beliefs about the parameters are summarized in their prior distribution. Then, new data values are collected experimentally (for example, patient survival in an oncology trial) and the probability distribution of these values leads to the likelihood function (the observed evidence on the parameters). The two elements are then combined, using Bayes' theorem, to produce the posterior distribution of the parameters — that is, the updated knowledge given the observed evidence. By contrast, frequentist methods rely solely on observed evidence for inferences, and typically do not formally take into account prior information.

Adaptive designs

In adaptive trial designs, interim data from a trial is used to modify and improve the study design, in a pre-planned manner and without undermining its validity or integrity. In the exploratory setting, an adaptive trial can assign a larger proportion of the enrolled subjects to the treatment arms that are performing well, drop arms that are performing poorly, and investigate a wider range of doses so as to more effectively select doses that are most likely to succeed in the confirmatory phase. In the confirmatory phase, adaptive design can facilitate the early identification of efficacious treatments, decisions to drop poorly performing trial arms, determining whether the trial should be terminated for futility and making sample-size adjustments at interim time points to ensure that the trial is adequately powered. In some cases, it might even be possible to enrich the patient population by altering the eligibility criteria at an interim time point.

Seamless designs

Such designs combine, in a single trial, the objectives that are traditionally addressed in separate trials. A seamless adaptive design addresses objectives normally achieved through separate trials using data from all trial stages, such as seamless adaptive Phase II/III trials.

Sample size re-estimation methods

These provide the flexibility to either increase or decrease the sample size at an interim point in the trial. This is important in cases in which there is uncertainty about between-subject variance in the response or uncertainty about the clinically meaningful effect size at which to power the trial. These methods allow the study to begin with a certain sample size that can be increased or decreased at an interim point, and even allow for an efficacy-stopping boundary.

in implementing these approaches. A PoC study in dyslipidaemia that illustrates the methods mentioned above is provided in Supplementary information S2 (box)).

Adaptive trial designs in early development. The core concept of adaptive trial design (also known as flexible design) is that it uses accumulating data to decide on how to modify aspects of the study mid-trial, in a pre-planned manner, without undermining the validity or integrity of the study[23–26]. Possible adaptations include adjustments to sample size, allocation of treatments, the addition or deletion of treatment arms, inclusion and exclusion criteria for the study population, adjusting statistical hypotheses (such as non-inferiority or superiority), and combining trials or treatment phases. Adaptive trials have the potential to translate into more ethical treatment of patients within trials, more efficient drug development and better use of available resources.

The standard approach to early development programmes is to separate the trials for PoC, dose ranging and dose selection. Adaptive designs offer several benefits over the standard approach. For example, a PoC trial can be combined with a dose-ranging trial (BOX 2). This approach has distinct advantages, in that it reduces start-up costs and the time between trials, and potentially increases statistical power and improves estimates of dose–response. Adaptive designs can also enable trialists to work with more candidate doses without increasing sample size. This is important to reduce risk of failure in confirmatory trials, where it has been estimated that, industry-wide, 45% of Phase III programmes do not have the optimum dose[3]. Adaptive dose-ranging studies are discussed further in BOX 3 (REFS 27,28).

There are a number of requirements for successful implementation of adaptive trial designs[23–26]**.** Drug responses need to be rapidly observable relative to accrual rate; alternatively, good longitudinal models can be used to forecast endpoints in time to adapt dose assignments for future subjects (assuming, of course, that the early measurements are good predictors of the late endpoint values). Adaptive trials also necessitate more up-front statistical work to model dose–response curves and to perform simulations — and many simulations are required to find the best combinations of sample size, the randomization ratio between placebo and drug, starting dose and number of doses. This in turn demands efficient programming to develop complex algorithms and fast computing platforms.

Confirmatory phase of development

The primary goals of a confirmatory clinical trial are to ensure that the diagnostic or therapeutic intervention causes less harm than good (safety) and to efficiently and confidently find the actual effect size on the chosen primary outcome(s) within the identified patient population (efficacy). Optimization of trial design during confirmatory development holds the promise of greater success rates, improved efficiency, better detection of safety signals, compressed timelines, smaller overall programmes and lower attrition rates. A number of novel approaches to confirmatory development that can contribute to fulfilling this promise are highlighted below.

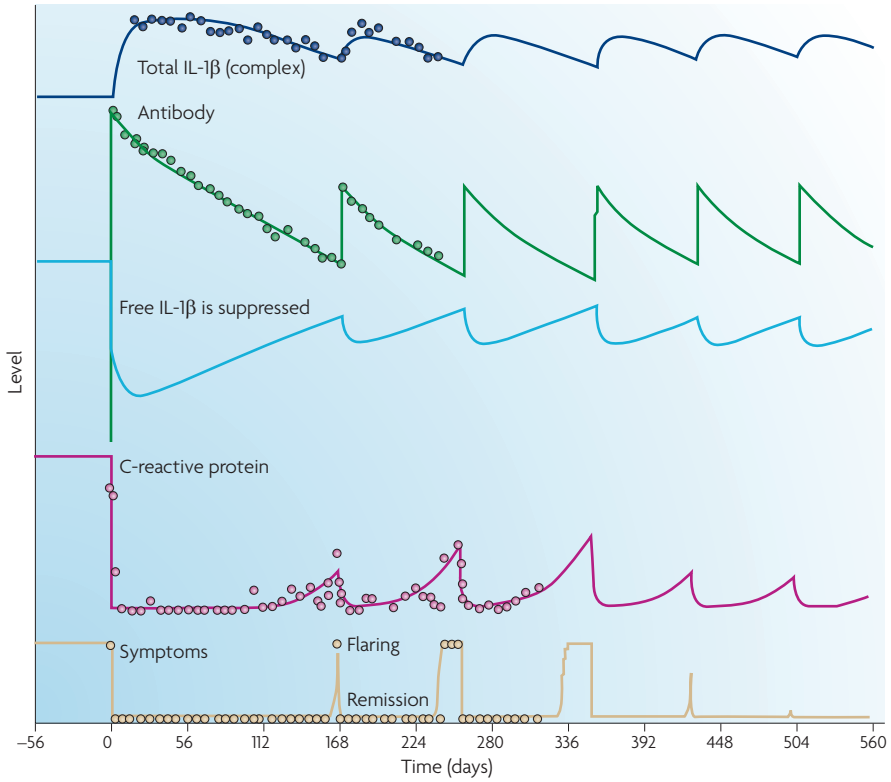


Figure 2 | **Dose selection in the development of a therapeutic for Muckle–Wells syndrome.** Muckle–Wells syndrome is a rare genetic disorder characterized by fever, urticaria, joint pain and malaise. A monoclonal antibody against interleukin-1β (IL-1 β), canakinumab, has been developed to treat this IL-1-dependent inflammatory disease. The antibody is delivered parenterally and binds to free IL-1β, driving it into the inactive complex and leading to remission of symptoms[21]. Total IL-1β, which represents mainly the inactive complex, increases after dosing and can be measured. By the laws of mass action, the free and active form of IL-1β, which cannot be measured, must decrease. However, the reduction in free IL-1β results in a decrease in markers of inflammation, including C-reactive protein (which can be measured), and a remission of clinical signs and symptoms of disease. The clinical data on these relationships can be captured in a mathematical model, shown in the figure, which is continuously adjusted in the light of new data. This framework simulation could then be used to propose a suitable dose and dosing regimen that would be predicted to produce a desired response for the majority of patients (for example, an 80% probability that 90% of patients will be flare-free for 2 months).

Seamless adaptive designs. Efficiency of the drug development process can be increased through the use of seamless adaptive designs, which aim to combine objectives traditionally addressed in separate trials into a single trial[25,29]. A specific example is the seamless adaptive Phase II/III design addressing objectives normally achieved through separate Phase II and III trials. These trials are confirmatory in nature, as opposed to seamless adaptive trials in early development, which are essentially exploratory. The first stage of a seamless adaptive Phase II/III trial might be similar to a late-Phase II trial, with a control group and several treatment groups (for example, different dose levels of the same treatment). Results are examined at the end of the first stage, and one or more of the treatment groups are selected to continue, along with the control group, into the trial's second stage. The final analysis comparing the selected group(s) with the control will use data from the continuing groups from both stages of the trial.

There are three key potential advantages of seamless adaptive designs: a reduction in the duration of the clinical development programme, by eliminating the time lag that traditionally occurs between Phase II and III trials; greater efficiency from the use of data from both stages, which might mean that fewer patients are required to obtain the same quality of information; and earlier acquisition of long-term safety data, gathered through continued follow-up of patients from the first stage (see Supplementary information S3 (figure))[25,29].

Not all drug development programmes will be candidates for these designs. Feasibility considerations for use of these designs include the length of follow-up time for the endpoint used for selection compared with duration of enrolment. Shorter follow-up will be more conducive to a seamless adaptive design, whereas a relatively long endpoint follow-up period will tend to militate against using such a design. Development programmes that do not involve complex treatment regimens might therefore be better suited to such designs. Drug supply and drug packaging will be expected to be more challenging in this setting.

A number of logistical and regulatory actions must be fulfilled to avoid compromising an adaptive trial. First, the actual algorithm for determining the adaptation to implement must be specified in advance. This is usually accomplished by creating a charter for the independent data monitoring committee charged with the responsibility of performing the unblinded interim analysis and communicating as appropriate with the sponsor. In addition, the sponsor must have developed in-house procedures to ensure that the algorithm is not transmitted throughout the company, and especially not to the study investigators.

To maintain trial integrity, the processes by which interim data are examined and selection decisions are made and implemented must be considered very carefully. Current conventions that restrict knowledge of interim results in ongoing trials should be respected to avoid compromising the interpretability of trial results. In some cases the decision being made at the selection point of a seamless design will be one for which sponsor perspective might be relevant and which has traditionally been a sponsor responsibility, raising the question of sponsor involvement in the monitoring process. A distinction is sometimes made between seamless adaptive designs that are inferentially seamless or operationally seamless. In inferentially seamless designs, which we describe here, the main analysis uses data from both stages of the trial. In operationally seamless designs, the final analysis only uses data from patients enrolled after the selection decision. This may allow a broader investigation of the first-stage data involving sponsor personnel and decreases concerns about trial integrity; in addition, traditional non-adaptive statistical methodology normally suffices. Such designs may maintain the advantage of reducing 'white space', while losing the efficiency that results from using

data accrued in stages. Regardless, operating procedures for the monitoring process in seamless designs must be carefully considered to ensure that the right expertise is applied to the decision, while limiting access to the accruing data as appropriate to maintain trial integrity.

Other considerations for adaptive designs include the endpoint used for selection. This need not be the same as the endpoint to be used in the main study analysis; if a good surrogate marker is available, this can be used and might enhance the efficiency of the seamless trial. Second, modelling and simulation will probably have a very important role in developing the specific details of seamless designs (for example, per-group sample sizes in the different stages, considered under various scenarios) to ensure that they are robust and efficient. Third, the final analysis must use statistical methodology that is appropriate for the design: 'naive' comparisons of control versus the selected treatment that do not account for the design will not be appropriate. Finally, the appropriateness of the design does not depend on any particular algorithm for choosing the patient group to be continued; it is not even necessary for a firm algorithm to be specified in advance, although the general principles that will govern the decision should be clear in advance.

Sample size re-estimation within a confirmatory trial (Phase III). Sample size re-estimation (SSR) provides a mechanism for appropriately using the information obtained during a confirmatory study to inform and adjust the necessary sample size going forward[30,31]. This process increases confidence that an appropriate sample size has been chosen to answer the primary study questions.

The standard approach used to power a confirmatory study is to first estimate the underlying treatment effect on the primary endpoint based on available prior information. The parameter δ denotes the true underlying difference between the treatment and control arms with respect to the primary endpoint. Even though the true value of δ is unknown, the trial investigators will usually have in mind a specific value, δ_{min}, which represents the smallest clinically important delta (SCID) for this clinical trial. Next, the trial designers will determine the sample size that can detect values of δ, based on prior information, that exceed the SCID with good power. The standard deviation σ (between subject variability) is a 'nuisance parameter' whose true value must be estimated in order to proceed with the sample size calculation.

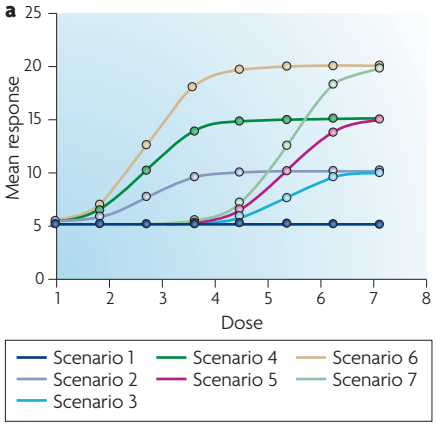
Box 2 | Case study: combining PoC and dose-ranging trials into a single adaptive trial

This example illustrates how a single adaptive trial can replace two standard trials — proof-of-concept (PoC) and dose-ranging — and that the combined trial has greater power than the standard PoC design, and is substantially better at estimating the dose–response curve.

The trial evaluated an analgesic drug to treat dental pain and tested seven doses of the drug. Several designs with different sample sizes, randomization ratios of drug to placebo and starting doses were simulated against several scenarios. Here, we describe one design with a sample size of 120 subjects (40 placebo, 80 drug). Bayesian adaptive trials were simulated over seven drug–response scenarios to enable comparisons with standard designs. Seven scenarios, which represent the gamut of probable dose–response curves were chosen as shown in panel **a** in the figure. In simulations, it was found that across all seven scenarios, a single adaptive trial can replace two standard trials (PoC and dose-ranging). The power of the trend test for PoC was always greater for the adaptive design, as shown in panel **b**. When there was a small dose–response effect (scenarios 2 and 3), the power of the adaptive design was about double that of the standard design. When the effect size was modest (scenarios 4 and 5), the power was increased to practically 100%. When effect sizes were large (scenarios 6 and 7), the power was almost 100% for both adaptive and standard designs.

For the same total sample size, the adaptive combined PoC–dose-finding trial is more efficient than the two standard trials in estimating the response at every dose (see panel **c**). The continuous curve shows the efficiency of the adaptive design relative to the standard dose-ranging design for scenario 7. Efficiency at each dose is defined as the ratio of the square of the estimation error of the standard design to the square of the estimation error of the adaptive design. The bars show the number of subjects allocated to each dose by the adaptive design. These results are computed by averaging the results of 1,000 simulations.The overall efficiency across all doses is greater by a factor of five, whereas for the sloping part of the dose response curve (doses 4, 5 and 6) the adaptive design is three times more efficient. In panel **d**, the adaptive combined PoC–dose-ranging trial with 60 subjects is as efficient in estimating the response at every dose as the two standard trials with a combined sample size of 120 subjects. It is also as powerful in testing for PoC.

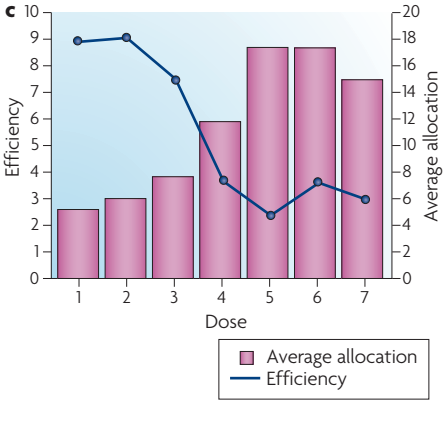
These results are true irrespective of which of the seven scenarios reflects the true dose–response curve. For all seven scenarios for the same sample size, the efficiency of the adaptive design was about five times that of the standard design over all doses. It was three times that of the standard design for estimating dose–response in the sloping part of the dose–response curve. Another way to think about this result is that for half the sample size, the adaptive design is as powerful and efficient as the standard approach with two trials.

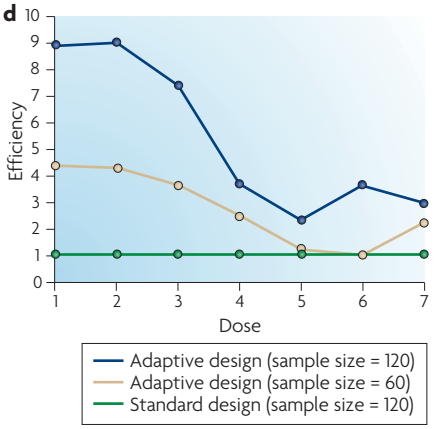


b

Scenario	Power of standard PoC trial*	Power of combined adaptive design
1	5	5
2	33	84
3	32	67
4	86	100
5	85	99
6	100	100
7	99	100

*Sample size = 30.

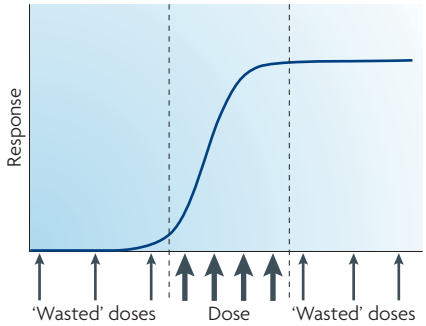




Box 3 | **Adaptive dose finding**

In an adaptive dose-finding study, the dose assignment(s) to the next subject, or next cohort of patients, is based on responses of previous subjects, and the dose assignment is chosen to maximize the information about the dose–response curve, according to some pre-defined objective metric (for example, variability in parameter estimates). In a traditional dose-finding trial, selecting a few doses may not adequately represent the dose–response relationship and many patients will be allocated to 'non-informative' doses (wasted doses), as shown in the figure. In adaptive dose-finding, the strategy is to initially include only a few patients on many doses to explore the dose–response, then to allocate the dose range of interest to more patients. This reduces the allocation of patients to non-informative doses[27,28]. Compared with fixed randomization, this approach has the ethical advantage that fewer subjects are assigned doses that are too high or too low. It can also avoid additional, separate trials that might be necessary when fixed dose-finding trials do not adequately define the dose range.

Adaptive dose-finding trials also require an infrastructure that allows the rapid communication of responses from trial sites to a central unblinded analysis centre and of adaptive dose assignments to the trial sites. Randomization software capable of rapidly computing dynamic allocation of doses to subjects is additionally mandated by adaptive trials because pre-specified randomization lists will not work. In addition, a flexible drug-supply process is required because demand for doses is not fixed in advance, but rather evolves as information on responses at various doses is gathered as the trial progresses.



The SCID can often be pre-specified from purely clinical arguments, whereas the actual effect size is unknown. Therefore, it is possible in principle to design a study with a fixed sample size that will have adequate power to detect the SCID, in the absence of adequate prior information about the actual effect size of the test agent. This is what statisticians envisaged when they created the fixed-sample methodology. However, this fixed sample methodology has several drawbacks. If the actual effect is substantially larger than the SCID, a smaller sample size would have sufficed to attain adequate power[32].

Sponsors will not often risk significant resources on trial sizes based on SCID assumptions that would lead to larger trials than the current 'best guess' about the actual effect size (BOX 4). Instead, a smaller trial corresponding to that best guess may be run; if that assumption is too optimistic, and the truth is an effect size closer to the SCID, the trial will be underpowered and therefore have a high chance of failure.

One approach to solving the problem of uncertainty about δ is to design and execute an additional number of exploratory trials (typically Phase II studies). These small Phase II studies are normally carried out to get a more precise estimate (or best guess) of the actual δ and σ so that the confirmatory study might be adequately powered. Each exploratory trial, although somewhat smaller than confirmatory trials, still requires significant resources to perform appropriately. Also, the inevitable start-up time and wind-down activities between trials have to be included when determining true programme efficiency and development timelines. This might therefore not be the most efficient way to proceed from the viewpoint of the entire clinical trial programme.

Box 4 | **Issues with the standard clinical development approach**

Issues with the standard approach to clinical development can be illustrated by considering a randomized clinical trial with the following assumptions. Based on available evidence from early-phase trials, it is estimated that $\sigma = 1$, that the anticipated effect size $\delta = 0.2$ and that the smallest clinically important delta (SCID) is 0.1. Rather than conservatively enrolling a sample size required to demonstrate the SCID (4,000 subjects), the sponsor appropriately powers the trial to detect the estimated larger δ (1,000 subjects). Now, suppose that the true underlying value of δ is 0.15. In that case, a sample size of 2,000 subjects would be required to adequately power the trial to detect this difference. The difficulty is, of course, that the true underlying value of δ is not known at the start of the trial. In this example, the 1,000-subject study would probably yield a non-significant result, as it is only powered to detect an effect size of 0.2, which is larger than the actual effect size of 0.15.

In this example, unless the 1,000-patient under-powered trial was repeated with a larger sample size, then a potentially efficacious treatment would be unnecessarily and unfortunately discarded. If the trial were to be repeated with the re-estimation of the actual effect size, then 2,000 patients would need to be enrolled, and the time and resources to perform the original trial (sometimes more than 3 years) would have been spent without much benefit other than gaining a more reliable estimate of the actual effect size in order to design the second trial. More importantly, the subjects for that study would have been put at unnecessary risk because the study had no real chance of being definitive.

Advantages of adaptive SSR in confirmatory trials. A more flexible approach to the fixed sample-size methodology is needed. By altering the sample size using interim data from the trial itself, this flexibility can be achieved without compromising the power or the false-positive rate of the trial (that is, the chance of making a false claim of efficacy for a treatment that is not efficacious). SSR should be considered in two situations: when there is significant uncertainty about σ; or when there is a substantial difference between the sample size resulting from using the SCID and the sample size the sponsor can justify on the basis of their best guess of the effect size[29].

SSR usually involves the choice of a suitable initial sample size, including one or more interim analyses at which the sample size will be re-assessed[30]. There are two distinct strategies — the group sequential strategy and the adaptive SSR strategy — for choosing the initial sample size, and then altering it on the basis of data obtained at various interim analysis time points. The group sequential strategy, which is also an adaptive design, begins with a large up-front sample size commitment and cuts back if the accruing data suggest that the large sample size is not needed. The adaptive SSR strategy proceeds in the opposite direction, starting out with a smaller initial sample size commitment but with the option to increase it should the accruing data suggest that such an increase is warranted[30–33] (BOX 5).

Extending the methodology to unknown σ. Although the group sequential and adaptive SSR methods were presented under the

assumption that the standard deviation σ is known, they apply equally for the case of unknown σ[30–32]. One can start out with an initial estimate of σ and corresponding sample-size estimate. Then, following an interim analysis, one can re-estimate this nuisance parameter, input the updated estimate into the equation and re-compute the sample size. An illustrative example is given in FIG. 3.

There are two ways to obtain the new sample size in the situation of unknown σ: blinded and unblinded. In the instance of blinded sample size re-estimation, the sponsor uses pooled data to estimate σ. This is permitted with no penalty to the analysis criteria (that is, alpha, or the probability of Type I (false positive) error). It is preferable that the sponsor pre-specifies how many times changes are to be made to the sample size, at what time points and how the new sample size will be calculated. Usually, this type of adjustment will not be permitted by regulatory authorities more than once.

For unblinded sample size re-estimation, the sponsor sets up a mechanism (possibly with the data monitoring committee of the trial) whereby the SSR is based on an unblinded estimate of variability (or statistical information) at the interim analysis. Sample size may be altered one or more times, but the maximum statistical information must be pre-specified.

If the sponsor agrees that there will be no early stopping for efficacy following an interim analysis, then no adjustment to the final analysis criteria is necessary. The data monitoring committee may monitor the data one or more times and adjust the sample size up or down based on the unblinded estimate of variability and attempt to reach the pre-specified maximum information.

When the sponsor pre-specifies the interim time points at which it is permissible to terminate early for efficacy, the criteria for each interim analysis must be pre-specified in a manner that controls the false-positive rate across the entire study. This will result in adjustment to the final analysis criterion if the study is not stopped early. Interim looks undertaken solely for administrative purposes, with no intention of stopping the trial in light of efficacy data, do not need to have defined criteria. The trial then proceeds until either it is terminated early for efficacy on the basis of the pre-defined criteria being reached, or until the planned maximum information (sample size or number of events) is reached.

Box 5 | **Group-sequential and adaptive designs for sample size re-estimation**

Group-sequential design. Suppose that the sponsor is unsure of the true value of δ, but nevertheless believes that it is larger than the smallest clinically important delta (SCID). In this case, a group-sequential design might be considered. Such a design is characterized by a maximum sample size, an interim monitoring strategy and a corresponding boundary for early stopping for efficacy. The maximum sample size is computed so that the study has adequate power to detect a value of δ that the sponsor believes represents a reasonable estimate of the efficacy of the experimental compound, provided this estimate is at least as large as the SCID. If the sponsor wishes to be very conservative about this estimate, the maximum sample size needed can be computed to have adequate power at the SCID itself. An up-front commitment is made to enrol patients up to this maximum sample size. However, if the true δ exceeds the SCID, the trial may terminate earlier with high probability by crossing an early stopping boundary at an interim analysis.

Returning to the example discussed in BOX 4, suppose that the sponsor decides to make an up-front commitment of 4,000 patients to the trial but intends to monitor the accruing data up to four times, after 1,000, 2,000, 3,000 and 4,000 patients become evaluable for the primary endpoint. The commitment of 4,000 patients ensures that the trial will have 88% power to detect a difference as small as δ = 0.1 (in this case the SCID). Although this is a rather large sample size to commit to the trial, the actual sample size is expected to be substantially smaller if the true δ is larger than the SCID. This is because at each of the four interim monitoring time points there is a chance of early termination and a declaration of statistical significance. At each interim analysis, a test for statistical significance using all available primary endpoint data would be performed, and the result would be compared with a properly determined early-stopping boundary value. The trial could be terminated the first time that a boundary is reached, with a valid claim that the experimental arm is more efficacious than the control arm.

However, sometimes a sponsor might not be willing to make such a large up-front commitment, particularly when the only currently available data on δ are from one or two small Phase II trials. The sponsor might feel more comfortable with a design that starts out with a smaller sample size of, say, 1,000 patients, with the opportunity to increase the sample size at an interim time point and after observing data from the trial. This is the motivation for the adaptive design below.

The adaptive design. The group-sequential design described above is characterized by pre-specifying a maximum sample size up-front and terminating earlier if the true δ is larger than anticipated. By contrast, an adaptive design pre-specifies a smaller initial sample size, but with the possibility of increasing the commitment after seeing some interim data from the trial. On the surface, this is similar to the usual practice of first running a small Phase II trial to obtain an idea about efficacy and safety and then following it up with a larger Phase III trial once the efficacy and safety of the compound have been established. There is, however, an important distinction between the conventional Phase II followed by Phase III strategy and the adaptive strategy outlined below.

In the conventional approach, the data from the Phase II trial are not combined with the data from the Phase III trial. The adaptive design, however, uses all the data from both stages for the final analysis. This can have important advantages both in terms of gaining additional statistical power, as well as shortening the drug development time. In our example, we stated that the SCID was 0.1. Supposing that the sponsor believes that the true δ = 0.2 — that is, twice as large as the SCID — if this is indeed the case, then a total sample size of 1,000 patients will have 89% power at a one-sided α level of 0.025. On this basis, the sponsor is prepared to make an initial investment of 1,000 patients to this trial. As an insurance policy, however, the sponsor intends to take an interim look at the accruing data at the mid-point of the trial, after 500 patients are evaluable for response. If the estimate of δ obtained from these 500 is smaller than the sponsor expected, then the sponsor might choose to increase the sample size to preserve the power of the trial.

Many different criteria can be used to decide whether an increase in sample size is warranted. A commonly used criterion is 'conditional power'. The conditional power at an interim analysis is the probability, given the observed data, that the experimental compound will demonstrate efficacy on completion of the trial. The conditional power computation requires specifying a value for δ. One can choose the value specified at the initial design stage or the value estimated from the interim data. In this example, we use the interim estimated value of δ for evaluating conditional power. The table below displays conditional power for various estimated values of δ at the interim time point, along with the total sample size needed to achieve 80% conditional power at the final analysis. The entries in the table assume that σ = 1.

Interim estimate (δ)	Conditional power without sample size increase	Total sample size needed to achieve 80% conditional power
0.2	95%	720 (sample size reduction)
0.175	86%	890 (sample size reduction)
0.15	72%	1,166 (sample size increase)
0.125	51%	1,757 (sample size increase)
0.1	30%	2,990 (sample size increase)

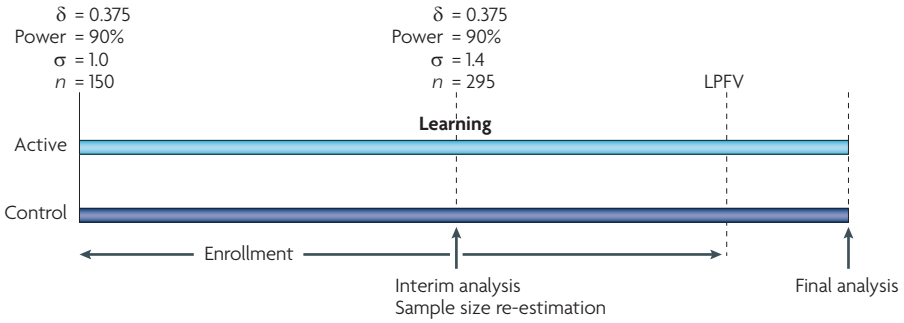


Figure 3 | **Re-estimating sample size while maintaining statistical power.** The figure illustrates a hypothetical example of a study in which sample size re-estimation due to uncertainty about the standard deviation σ led to an increase in sample size to ensure 90% power was maintained. At the beginning of the trial, the planned sample size was estimated at 150 patients based on a standard deviation of 1.0. At the interim analysis, the actual standard deviation was 1.4. Even though the effect size (δ) was as originally predicted, an increase in sample size to 295 patients would be required to maintain 90% power. Without the sample size re-estimation, the power at the final analysis would only be 64% and there would be much greater risk of a failed trial. LPFV, last patient first visit.

Tackling challenges of new trial designs

Because they are so flexible, these new trial designs require significant statistical analyses, simulations and logistical considerations to verify their operating characteristics, and therefore tend to require more time for the planning and protocol development phase. Regulatory agencies and Institutional Review Boards also need to approve the design format for interim analysis, and these discussions can sometimes take considerable time. Such time considerations can lead a company to follow the traditional route to clinical development, without fully appreciating the advantages that adaptive designs can eventually bring in terms of time and cost savings, and probability of success.

As described above, adaptive designs further require the following: quickly observable responses relative to the patient accrual rate or good longitudinal forecasting models; efficient design and implementation software and fast computing platforms; an infrastructure that facilitates rapid communication across trial sites to the central unblinded analysis centre and rapid communication of dose assignments to trial sites; and a flexible drug-supply process. Appropriate models, which reliably characterize the longitudinal behaviour of clinical endpoints, or the relationship between biomarkers and endpoints, are also crucial to the success of the modern clinical development paradigm discussed here. Because model assumptions often need to be checked — and at times revised — after data have been observed, an intriguing possibility is to use 'adaptive modelling' approaches. This is a topic for further research, and is beyond the scope for this paper.

Maximizing the use of all potential prior information requires greater collaboration across functional silos in organizations to avoid compartmentalization of data. In practice, the inclusion of a broader sample of datasets can be difficult because of a lack of common data standards. These problems are compounded by competitive hurdles to sharing what is considered proprietary information about novel therapies without PoC, which inhibits the exchange of data. Overcoming internal resistance and aversion to change also represents a major hurdle for incorporating the prospective use of novel trial designs and methodologies, and modelling and simulation, into clinical development programmes.

A key challenge for the implementation of tools and techniques which advance the quality, timeliness and efficiency of drug development is the ability to work across disciplines and amongst stakeholders to understand how and when to apply these solutions. To address this challenge, we make the following recommendations. First, a common vocabulary and a common understanding of the value of modern trial designs to all stakeholders needs to be defined and disseminated. Second, at the same time, guidelines and case studies for assessing situations in which tools should be applied, as well as for those scenarios when they should not be utilized, should be developed and disseminated. Third, there is a need to create a methodology for dialogue with regulatory authorities to facilitate discussion of clinical strategies which utilize these tools and address potential constraints and issues. Finally, it will be crucial to identify specific solutions to address all mindset obstacles and factual objections that inhibit the adoption of modern tools and adaptive study designs.

John Orloff, Jose Pinheiro, Michael Branson, Paul Gallo and Donald Stanski are at Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, New Jersey 07936, USA.

Frank L. Douglas is at the Austen BioInnovation Institute, 1 South Main Street, Suite 401, Akron, Ohio 44308, USA.

Susan Levinson is at The Strategic Choice LLC, 6 East Cove Lane, Morristown, New Jersey 07960, USA.

Pravin Chaturvedi is at Napo Pharmaceuticals, 250 East Grand Avenue, Suite 70, South San Francisco, California 94080, USA.

Ene Ette is at Anoixis Corporation, 214 North Main Street, Suite 104, Natick, Massachusetts 01760, USA.

Gigi Hirsch, Nitin Patel, Sameer Sabir, Stacy Springs and Howard Golub are at the Center for Biomedical Innovation, Massachusetts Institute of Technology, 77 Massachusetts Avenue (E19-611), Cambridge, Massachusetts 02139-4307, USA.

Cyrus Mehta is at Cytel Statistical Software and Services, 675 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

Matthias R. Evers, Edd Fleming, Navjot Singh and Tony Tramontin are at McKinsey & Company, Inc., Am Sandtorkai 77, 20457 Hamburg, Germany.

Correspondence to J. O.
e-mail: john.orloff@novartis.com

doi:10.1038/nrd3025
Published online 9 October 2009

1. Booth, B & Zemmel, R. Prospects for productivity. *Nature Rev. Drug Discov.* **3**, 451–456 (2004).
2. Gilbert, J., Henske, P. & Singh, A. Rebuilding big pharma's business model. *In Vivo* **21**, 1 – 4 (2003).
3. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nature Rev. Drug Discov.* **3**, 711–716 (2004).
4. Budget US Govt., App., FY 1993–2003.
5. *Parexel's Pharmaceutical R&D Statistical Sourcebook 2002/2003* (Parexel International, Waltham, USA, 2003).
6. Adams, C. P. & Brantner, V. V. Spending on new drug development. *Health Econ.* 26 Feb 2009 (doi: 10.1002/hec.1454).
7. DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22**, 151 – 185 (2003).
8. Adams, C. & Brantner, V. V. Estimating the cost of new drug development: is it really $802 million? *Health Aff.* **2**, 420–428 (2006).
9. *Pharmaceutical R&D Factbook 2007* (CMR International, London, UK, 2007).
10. *KMR General Metrics Study* (KMR Group, Chicago, USA, 2007).
11. Sheiner, L. B. & Steimer, J.-L. Pharmacokinetic/ pharmacodynamic modeling in drug development. *Ann. Rev. Pharmacol. Toxicol.* **40**, 67–95 (2000).
12. Lalonde, R. L. *et al.* Model-based drug development. *Clin. Pharmacol. Ther.* **82**, 21–32 (2007).
13. Breimer, D. D. & Danhof, M. Relevance of the application of pharmacokinetic–pharmacodynamic modeling concepts in drug development. The "Wooden Shoe" paradigm. *Clin. Pharmacokinet.* **32**, 259–267 (1997).
14. Danhof, M., Alvan, G., Dahl, S. G., Kuhlmann, J. & Paintaud, G. Mechanism-based pharmacokinetic/ pharmacodynamic modeling — a new classification of biomarkers. *Pharm. Res.* **22**, 1432–1437 (2005).
15. Holford, N. H. G., Kimko, H. C., Monteleone, J. P. R. & Peck, C. C. Simulation of clinical trials. *Annu. Rev. Pharmacol. Toxicol.* **40**, 209–234 (2000).
16. Miller, R. *et al.* How modeling and simulation have enhanced decision making in new drug development. *J. Pharmacokinet. Pharmacodyn.* **32**, 185–197 (2005).
17. Berry, D. A. Bayesian statistics. *Med. Decis. Making* **26**, 429–430 (2006).
18. Müller, P., Berry, D. A., Grieve, A. P., Smith, M. & Krams, M. Simulation-based sequential Bayesian design. *J. Stat. Plan. Inference* **137**, 3140–3150 (2007).

19. Goggin, T. *et al.* Modeling and simulation of clinical trials: an industrial perspective. In *Simulation for Designing of Clinical Trials* (eds Kimko, H. C. & Duffull, S. B.) 227–224 (Marcel Dekker, New York, USA, 2002).
20. Reigner, B. G. *et al.* An evaluation of the integration of pharmacokinetic and pharmacodynamic principles in clinical drug development. Experience within Hoffman La Roche. *Clin. Pharmacokinet.* **33**, 142–152 (1997).
21. Lachmann, H. J. *et al.* Use of canakinumab in the cryopyrin-associated periodic syndrome. *N. Engl. J. Med.* **360**, 2416–2425 (2009).
22. Berry, D. A. Bayesian clinical trials. *Nature Rev. Drug Discov.* **5**, 27–36 (2006).
23. Berry, D. A. Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. *Clin. Trials* **2**, 295–300 (2005).
24. Berry, D. A. Adaptive trial design. *Clin. Adv. Hematol. Oncol.* **5**, 522–524 (2007).
25. Krams, M. *et al.* Adaptive designs in clinical drug development: opportunities, challenges, and scope reflections following PhRMA's November 2006 workshop. *J. Biopharm. Stat.* **17**, 957–964 (2007).
26. Gallo, P. *et al.* Adaptive designs in clinical drug development — an executive summary of the PhRMA Working Group. *J. Biopharm. Stat.* **16**, 275–283 (2006).
27. Bornkamp, B. *et al.* Innovative approaches for designing and analyzing adaptive dose-ranging trials. *J. Biopharm. Stat.* **17**, 965–995 (2007).
28. Weir, C. J., Spiegelhalter, D. J. & Grieve, A. P. Flexible design and efficient implementation of adaptive dose-finding studies. *J. Biopharm. Stat.* **17**, 1033–1050 (2007).
29. Bretz, F., Schmidli, H., König, F., Racine, A. & Maurer, W. Confirmatory seamless Phase II/III clinical trials with hypotheses selection at interim: general concepts. *Biom. J.* **48**, 623–634 (2006).
30. Mehta, C. R. & Patel, N. R. Adaptive, group sequential and decision theoretic approaches to sample size determination. *Stat. Med.* **25**, 3250–3269 (2006).
31. Chuang-Stein, C., Anderson, K., Gallo, P. & Collins, S. Sample size reestimation: a review and recommendations. *Drug Inf. J.* **40**, 475–484 (2006).
32. Golub, H. L. The need for more efficient clinical trials. *Stat. Med.* **25**, 3231–3235 (2006).
33. Tsiatis, A. A. & Mehta, C. On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika* **90**, 367–378 (2003).

Acknowledgements
A special acknowledgment to W. Dere (Amgen), S. Cummings (UCSF), A. Lee (Pfizer) and E. Berndt (MIT-CBI) for significant contributions to discussions leading to this manuscript. Also, many thanks to the McKinsey Trial Design Team for their support (M. E., E. F., N. S. and T. T.).

Competing interests statement
The authors declare competing financial interests: see web version for details.

DATABASES
OMIM:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Muckle–Wells syndrome | type 2 diabetes

SUPPLEMENTARY INFORMATION
See online article: S1 (box) | S2 (box) | S3 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF