

Statistical refocusing in the design of Phase II trials offers promise of increased R&D productivity

Stacy R. Lindborg, Charles C. Persinger, Andreas Sashegyi, Craig Mallinckrodt and Stephen J. Ruberg

The process of drug research and development (R&D) is expensive and time-consuming, and productivity in terms of the number of new drugs resulting from a given level of investment in R&D has been declining for decades, as highlighted in a recent article (Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Rev. Drug Discov.* **11**, 191–200 (2012))¹. Despite many calls for innovation to the drug development process, few substantively different proposals have been articulated or used. Here, we propose a simple and implementable means to increase productivity by focusing on decisions that are under the industry's control — the choices of statistical parameters in Phase II trial design.

Like many drug development decisions, decisions about the statistical design of Phase II trials are typically made at the level of an individual drug development programme, often with cost and time as key considerations. We use a different starting point for such decisions by focusing on risk, as embodied in the probability of making a false positive (FP; advancing a compound that is not a viable drug) or false negative (FN; terminating a useful compound) decision (FIG. 1a). We use a systems approach to put these decisions in the context of an entire R&D portfolio, rather than optimizing at the trial or compound level. We developed a model to assist with this evaluation, and the output suggests that R&D productivity for

companies with a portfolio of compounds could be improved by alterations to current practice.

Rigorous modelling and quantification of strategic and business decisions required during clinical development can supplement intuition and judgement, reveal levels of uncertainty in outcomes, and foster insights into interrelationships and implications. By extending work done by Paul and colleagues², which examined the relative importance of cost, cycle time and success rates across eight phases in preclinical and clinical development on overall R&D productivity, we created a model called the 'Strategic R&D portfolio model' (FIG. 1b). Using this model, we can provide detailed analyses based on industry-wide data to address company-specific questions regarding how to most effectively improve productivity.

The work by Paul and colleagues highlighted the importance of several parameters that are influenced by Phase II statistical design decisions: the probability of technical success ($p(\text{TS})$), the cycle time and the cost of Phase II and Phase III trials. The primary statistical design decision centres on the required degree of confidence in the results. Put another way, what is the optimum probability of making the 'right' decision or,

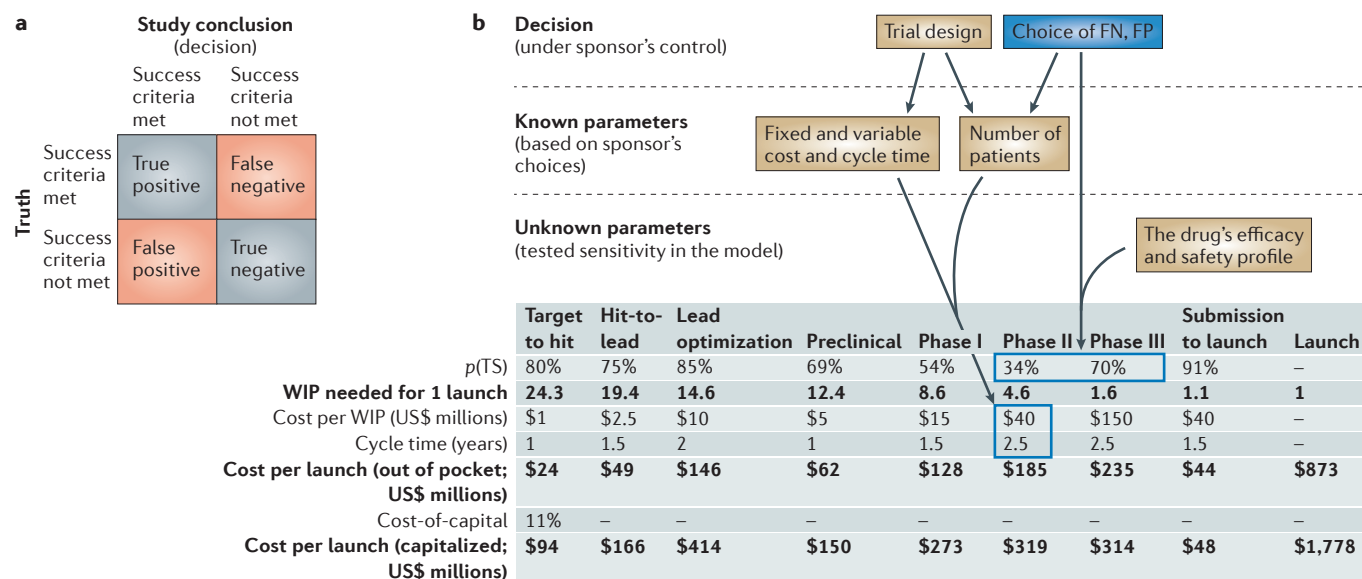


Figure 1 | **The strategic R&D portfolio model.** **a** | The figure outlines the four possible conclusions that can be drawn about a clinical trial. Two conclusions match the truth and result in a true negative or positive decision, while the other two conclusions are false. A key goal in the statistical design of a clinical trial is to control the likelihood of a false decision to an acceptable level. Historically, in Phase II, 80% power and $\alpha=0.05$ have been viewed as acceptable levels to control false negative (FN) and false positive (FP) rates, respectively, and are commonly used in the context of a hypothesis testing framework. **b** | Our model, which was adapted from work by Paul and

colleagues² and is used to calculate the cost per drug launch, can be used to identify the implications of changes to the input parameters of probability of technical success ($p(\text{TS})$), work in process (WIP) needed for one drug launch, cost and cycle time. As highlighted in this figure, the choice of power and α level, along with other drug and trial parameters, defines the values for Phase II and Phase III $p(\text{TS})$ and Phase II cost and cycle time. By defining the relationships among the parameters, the implications of using the most appropriate FP and FN (determines the power) levels can be appreciated. See Supplementary information S1 (box) for details.

conversely, what are the optimum FP and FN findings? It is important to maximize the probability of the right decision, thereby minimizing FP and FN rates.

The FP and FN errors are equivalent to α and β in the simple hypothesis testing paradigm, but can be generalized to encompass more complex success factors that go beyond merely testing whether the mean efficacy response exceeds a certain threshold. For example, success could be judged on a composite measure of efficacy and safety, which would entail a more complex evaluation of the FP or FN probabilities. Typically, Phase II studies are designed with one primary objective that is often evaluated by a simple hypothesis test with 80% power ($\beta = \text{FN} = 20\%$) and α of 0.05 (FP = 5%). However, this analysis supports the consideration of other alternatives.

The strategic R&D portfolio model

Our model (FIG. 1b) incorporates three input parameters: the probability that a molecule makes it through the phase (that is, $p(\text{TS})$), the cycle time (in years) that it takes a molecule to make it through the phase, and the cost (in US dollars) that it takes to progress a molecule through the phase. We then adjusted the decision parameters — FP and FN — to determine the work in process (WIP) needed to launch one drug and the associated costs. We used a base case of a typical clinical development plan that has one Phase II trial, which is assumed to be a dose–response trial of three doses and placebo, followed by two Phase III trials done in parallel.

The productivity measure was calculated in the model with FP and FN values ranging from 0.01 to 0.50, assuming a 50% probability that the drug was efficacious (FIG. 2; see [Supplementary information S1](#) (box) for details). The analysis showed that, for the base case tested, the most productive choices were in the elliptical region with an FP between 0.15 and 0.35 and an FN between 0.05 and 0.15. These rates differ appreciably from the typical choice of FN = 0.2 (or 80% power) and FP = 0.05 (FIG. 2). We extended results by varying the probability that the drug was efficacious from 30% to 70%. The optimal region for FP and FN rates did not change appreciably (results not shown in FIG. 2).

Given that the optimal region differs from the historical choices, it is important to understand why changes to the FP rate and FN rate affect R&D productivity. Our simulations reveal that a move from the historical choices (80% power = $1 - \text{FN}$; FP = 5%) to the highlighted region shown in FIG. 2 could

result in a decrease of approximately 6–7% in cost per launch (out of pocket), a change that is completely under the sponsor's control. Using the capitalized cost per launch, the productive region would shift even more towards higher power (lower FN) and relaxing (increasing) FN. A move from the historical choices of power and FN to the recommended region would result in a decrease of approximately 10% in capitalized cost per launch, resulting in higher productivity.

Implications

Our analysis indicates that the choice of FP and FN rates for Phase II trials — which is completely under the control of sponsors — can appreciably influence R&D productivity. Importantly, in the simple case evaluated in our simulations, and assuming implementation on an existing Phase II portfolio, the change would probably not require increases in Phase II costs or time. More Phase II projects would succeed, causing more projects to enter Phase III and increase Phase III spending. However, more product launches would result, which would offset the Phase III spending.

At steady state, these recommendations could reduce cost because fewer projects are needed to achieve the same level of output. However, the more likely approach would be to keep the same level of overall spending but to generate more product launches (and value to patients and shareholders). For a large pharmaceutical or biotechnology company spending more than US\$2 billion per year on clinical R&D and operating to produce two new molecular entity (NME) launches per year, a 6–10% productivity improvement could result in the steady-state increase of one or two new drug launches every 10 to 15 years without any change in overall spending.

Based on our modelling, we advocate a more strategic approach to the problem of optimizing a large clinical portfolio; that is, to choose appropriate FP and FN levels and control the risk of making the wrong decision. The lost revenue (that is, opportunity cost) stemming from terminating a drug that is in truth effective is typically much greater than the cost of advancing an ineffective molecule into Phase III. Therefore, intuitively it makes sense that the optimum FN rate

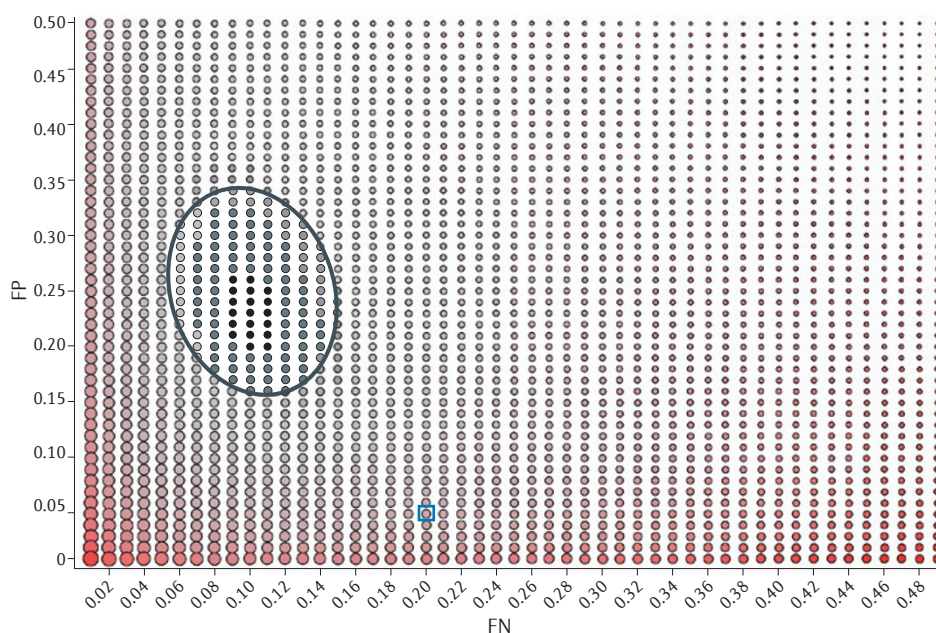


Figure 2 | Optimal false positive and false negative rates in Phase II trials to achieve the highest productivity across an R&D portfolio. Using different choices for false positive (FP) and false negative (FN) results in different outputs, as generated from multiple simulation runs associated with the model and assumptions introduced in FIG. 1. Each dot represents a specific combination associated with FP and FN values. The size of the dot represents the number of patients required to achieve the level of FP and FN (risk), and the colour of the dot represents the calculated productivity metric (red = higher cost per new molecular entity (NME) and black = lower cost per NME). The area of greatest productivity is contained in the ellipsoid ($0.15 < \text{FP} < 0.35$ and $0.06 < \text{FN} < 0.15$ or $0.85 < \text{power} < 0.94$). This range of values differs from the traditional choice of FP = 0.05 and power = 0.80 (shown by the dot outlined by the blue box). This graph assumes a 50% probability that the drug is efficacious. See [Supplementary information S1](#) (box) for details.

should be lower than the optimum FP rate, as FN mistakes are more costly. Although this is common sense, it has not been common practice. In fact, the opposite has been true, perhaps owing to convention in Phase III. Regardless, decisions on allocating such risk need to consider market size, competitors in development and in the market, and the probability that the drug is effective, as this in turn defines the opportunity for FN versus FP decisions.

The optimal region of FP and FN levels was fairly consistent over a wide range of values for the input parameters. This provides greater confidence in the generalizability of our conclusions. Nevertheless, we recognize that judgement is required for making decisions about any specific Phase II plan. For example, the conclusions would change if the cost of a failed Phase III study were much higher than were modelled (for example, from more expensive Phase III trials, substantial additional costs in Phase III related to marketing or manufacturing capital expenses, or negative

reputational impact from failed Phase III studies). In these situations, guarding against FPs may be worth the increased expense in Phase II or the sponsor may need to offset guarding against FPs by accepting a higher probability of FNs. Still, keeping in mind the nature of a strategy and the advantage that accumulates through adherence to a consistent pattern of behaviour³, the starting point for any Phase II programme discussion should be based on models and simulations tailored to the specific situation.

In conclusion, we found that the FP and FN levels commonly used with study protocol design (simple hypothesis tests with $\alpha = 0.05$ and $\beta = 0.2$) are far from optimal for Phase II in the scenarios simulated in our model. Our analysis indicates that switching these values could increase productivity by at least 6%. This improvement in productivity would not require more study patients, higher costs or increased time. For a large pharmaceutical or biotechnology company, the change could result in one or two NMEs per decade on the

market, simply by changing the statistical lens used during Phase II study design and analysis.

Stacy R. Lindborg is at Biogen Idec,
14 Cambridge Center, Cambridge,
Massachusetts 02142, USA.

Charles C. Persinger, Andreas Sashegyi,
Craig Mallinckrodt and Stephen J. Ruberg are at
Eli Lilly and Company, Lilly Corporate Center,
Indianapolis, Indiana 46285, USA.

Correspondence to S.R.L.
e-mail: bayespower@gmail.com

doi: 10.1038/nrd3681-c1

1. Scannell, J. W. *et al.* Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Rev. Drug Discov.* **11**, 191–200 (2012).
2. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Rev. Drug Discov.* **9**, 203–214 (2010).
3. Pisano, G. *Creating an R&D Strategy*. Working Paper 12–095 (Harvard Business School, 2012).

Competing interests statement

S.R.L., C.C.P., A.S., C.M. and S.J.R. own stock of the companies that they are employed by.

SUPPLEMENTARY INFORMATION

See online article: [S1](#) (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF