How to improve R&D productivity: the pharmaceutical industry’s grand challenge

Steven M. Paul, Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg and Aaron L. Schacht

Abstract | The pharmaceutical industry is under growing pressure from a range of environmental issues, including major losses of revenue owing to patent expired, increasingly cost-constrained healthcare systems and more demanding regulatory requirements. In our view, the key to tackling the challenges such issues pose to both the future viability of the pharmaceutical industry and advances in healthcare is to substantially increase the number and quality of innovative, cost-effective new medicines, without incurring unsustainable R&D costs. However, it is widely acknowledged that trends in industry R&D productivity have been moving in the opposite direction for a number of years. Here, we present a detailed analysis based on comprehensive, recent, industry-wide data to identify the relative contributions of each of the steps in the drug discovery and development process to overall R&D productivity. We then propose specific strategies that could have the most substantial impact in improving R&D productivity.

The pharmaceutical industry is facing unprecedented challenges to its business model. Experienced observers and industry analysts have even predicted its imminent demise1-3. Over the past decade, serious concerns about the industry's integrity and transparency — for example, around drug safety and efficacy — have been raised, compromising the industry's image, and resulting in increased regulatory scrutiny4-6. This erosion in confidence in the industry and its products has resonated poorly with patients, health-care professionals, payers and shareholders. Indeed, the industry’s price/earnings ratio, a measure of the current valuation of the industry, has decreased below that of the S&P 500 index and has remained more or less flat, as have share prices for the past 7 years.

The industry's profitability and growth prospects are also under pressure as healthcare budgets become increasingly strained. Generic drugs, although clearly helping to keep drug prices in check, are currently approaching 70% of all prescriptions written in the United States6. Moreover, key patent expirations between 2010–2014 have been estimated to put more than US$209 billion in annual drug sales at risk, resulting in $113 billion of sales being lost to generic substitution7. Indeed, for every dollar lost in declining product revenues due to patent expirations by 2012, it has been estimated that large-cap pharmaceutical companies will only be able to replace on average 26 cents with new product revenues8.

Simply stated, without a dramatic increase in R&D productivity, today’s pharmaceutical industry cannot sustain sufficient innovation to replace the loss of revenues due to patent expirations for successful products. A key aspect of this problem is the decreasing number of truly innovative new medicines approved by the US Food and Drug Administration (FDA) and other major regulatory bodies around the world over the past 5 years (in which 50% fewer new molecular entities (NMEs) were approved compared with the previous 5 years)9. In 2007, for example, only 19 NMEs (including biologics) were approved by the FDA, the fewest number of NMEs approved since 1983, and the number rose only slightly to 21 in 2008. Of the 21 new drugs approved by the FDA in 2008, only 6 were developed by the 15 largest pharmaceutical companies and only 29% would be considered 'first-in-class' medicines. In 2009, 24 new drugs were approved, 10 of which were developed by large pharmaceutical companies and only 17% of which could be considered first-in-class. Some have argued that the number of approved 'mechanistically
innovative' and first-in-class NMEs have remained stable at about 5–6 per year. However, the number of potential revenue-generating drugs (innovative or otherwise) as a percentage of R&D expenditures has undeniably fallen sharply.

With an estimated $50 billion in collective annual R&D spending by the large pharmaceutical companies, and appropriate allocation over time to the successful discovery and development of NMEs, the average cost for these companies to bring an NME to market is now estimated to be approximately $1.8 billion (see below for details underlying this estimate), and is rising rapidly.

Moreover, there is little evidence that the average costs of successfully launching an NME vary significantly between large pharmaceutical or small biotechnology companies.

Although R&D productivity has been declining for a number of years, the unprecedented combination of reduced R&D output in the form of successfully launched truly innovative NMEs, coupled with diminishing market exclusivity for recently launched new medicines and the huge loss of revenues owing to generic competition over the next decade, suggest that we may be moving closer to a pharmaceutical ‘ice age’ and the potential extinction of the industry, at least as it exists today. Although this might be welcomed by the industry’s critics, the impact on the health and well-being of patients owing to delayed or even lost opportunities to introduce the next generation of innovative medicines could be devastating. In this regard, we underscore the findings of Lichtenberg on the effects of medical innovation (including controls for the impact of obesity and income), which indicate that ~40% of the 2-year increase in life expectancy measured from 1986–2000 can be attributed to the introduction and use of new drugs. It took approximately 3 years for NME launches to have their maximal impact on longevity — this effect was not observed for non-NME (older) drugs. One can only speculate as to the impact on longevity and quality of life that new drugs now in clinical development for cancer and Alzheimer’s disease might have. Without these new medicines, and given the rise in diseases such as diabetes and childhood obesity, it is possible that life expectancy may actually decrease over time.

Among all the challenges faced by the pharmaceutical industry, we argue that improving R&D productivity remains the most important. The environmental factors that are reducing the industry’s profitability can only be mitigated by substantially and sustainably increasing the number and quality of innovative, as well as cost-effective, new medicines; but only if accomplished at reasonable R&D costs. So, the key questions are where, how and by how much can R&D productivity be improved? Here, we present a detailed analysis of R&D productivity by first defining and modelling the essential elements of contemporary drug discovery and development that account for the current cost of a new medicine, and discuss the rate-limiting steps of the R&D process that are contributing to reduced R&D productivity. We then propose, and illustrate, ways to improve these factors.

How do we define R&D productivity?

R&D productivity can be simply defined as the relationship between the value (medical and commercial) created by a new medicine (considered here to be an NME) and the investments required to generate that medicine. However, R&D productivity can in our view best be elaborated in two important dimensions: inputs leading to outputs, or R&D efficiency; and outputs leading to outcomes, or R&D effectiveness (FIG. 1).

R&D efficiency represents the ability of an R&D system to translate inputs (for example, ideas, investments, effort) into defined outputs (for example, internal milestones that represent resolved uncertainty for a given project or product launches), generally over a defined period of time. If launching (gaining regulatory approval and commercializing) an NME is the desired output, how can this be achieved with greater efficiency (that is, at a lower cost)?

R&D effectiveness can be defined as the ability of the R&D system to produce outputs with certain intended and desired qualities (for example, medical value to patients, physicians and payers, and substantial commercial value). Thus, R&D productivity can be viewed as an aggregate representation of both the efficiency and effectiveness of the drug discovery and development process; the goal of a highly productive R&D system is to efficiently translate inputs into the most desired and valuable outputs. For a more detailed description of these definitions, see Supplementary information S1 (box).

With this definition of R&D productivity in mind, we have further adapted a productivity relationship or ‘pharmaceutical value equation’, which includes the key elements that determine both the efficiency and effectiveness of the drug discovery and development process for any given pipeline (see equation 1).

\[
P = \frac{\text{WIP} \times p(\text{TS}) \times V}{\text{CT} \times C}
\]

R&D productivity (P) can be viewed as a function of the elements comprising the numerator — the amount of scientific and clinical research being conducted simultaneously, designated here as the work in process (WIP), the probability of technical success (p(\text{TS})) and the value (V) — divided by the elements in the denominator, the cycle time (CT) and cost (C). Each of these parameters can be conceptualized and analyzed on a per project basis (for example, a single drug candidate or WIP = 1) or collectively as a larger portfolio or pipeline of projects or drug candidates. In general, increasing the numerator relative to the denominator will increase productivity and vice versa. Thus, if one could increase the p(\text{TS}) (that is, reduce attrition) for any given drug candidate or ideally for a portfolio of drug candidates at a given phase of development, \( P \) would increase accordingly. Similarly, for any given level of R&D investment, substantially reducing \( CT \) or lowering \( C \) (such as unit costs) would increase \( P \).

However, most of the elements comprising equation 1 are inextricably linked to one another and changing one element can often adversely or beneficially affect
This is the out-of-pocket cost corrected for cost of capital, and is used to evaluate the real economic value of a project. It is a measure of the incremental costs required to fund and execute a project above what would be required for a similar project in a different setting.

Out-of-pocket cost
This is the total cost required to expect one drug launch, taking into account attrition, but not the cost of patents. This is the cost that investors must consider when deciding to invest in a project.

Cost of capital
This is the annual rate of return expected by investors based on the level of risk of the investment.
20 and 15 respectively (Fig. 2). We will discuss the need for sufficient discovery investments and output (WIP) to achieve the level of drug candidates necessary below. In this model, in the absence of sufficient acquisition of drug candidates, especially late-phase compounds, achieved as one-off in-license deals or through mergers and acquisitions (M&As), most companies are simply unable to achieve (or afford) the numbers of compounds distributed across the phases of discovery and development they require to achieve their goals for new NMEs launched without a substantial increase in productivity.

Encouragingly, recent benchmark data on Phase I WIP across the industry indicate that most companies have begun to substantially increase investments in the earlier stages of drug discovery; this is reflected by the number of candidates entering Phase I trials, which has increased significantly\(^{1,2,3}\). However, based on the benchmark data, for most companies, the number of NMEs entering clinical development and progressing to Phase II and III are still insufficient to achieve 2–5 launches per year\(^4\); this reflects many years of operating at WIP levels below what would be required in the earlier stages of drug discovery and development. Thus, inevitable pipeline gaps will arise (as they have) and given the CT of the process (Fig. 2), such gaps cannot be filled quickly through traditional means.

Finally, we suggest that based on this model, many companies would find that their R&D operating expenses are not appropriately distributed across the various phases of drug discovery and development. Too many resources are often applied to late-stage development of drug candidates with relatively low \(p(TS)\) and/or post-launch support of marketed products. This may be the root cause of the current drought of new medicines and the business challenges most companies are experiencing.

### Key areas for improving R&D productivity

Using our model (Fig. 2, Supplementary information S2 box) and starting from a baseline value for the estimated capitalized cost of a single NME of ~$1.78 billion, we can investigate which parameters contributing to this cost are the most important. To achieve this, we have varied the parameters \(p(TS)\), CT and C for different phases of the overall process across a realistic range of possibilities (reasonable estimates of industry highs and lows for each parameter) to identify parameters for which changes would have the greatest impact on R&D efficiency, and the extent of the impact in each case (Fig. 3).

As is evident from Fig. 3, attrition — defined as \(1 - p(TS)\) — in the clinical phases of development (especially Phase II and III) remains the most important

---

**Figure 2** | R&D model yielding costs to successfully discover and develop a single new molecular entity. The model defines the distinct phases of drug discovery and development from the initial stage of target-to-hit to the final stage, launch. The model is based on a set of industry-appropriate R&D assumptions (industry benchmarks and data from Eli Lilly and Company) defining the performance of the R&D process at each stage of development (see Supplementary information S2 box for details). R&D parameters include: the probability of successful transition from one stage to the next \(p(TS)\), the phase cost for each project, the cycle time required to progress through each stage of development and the cost of capital, reflecting the returns required by shareholders to use their money during the lengthy R&D process. With these inputs (darker shaded boxes), the model calculates the number of assets (work in process, WIP) needed in each stage of development to achieve one new molecular entity (NME) launch. Based on the assumptions for success rate, cycle time and cost, the model further calculates the 'out of pocket' cost per phase as well as the total cost to achieve one NME launch per year (US$1.78 billion). Lighter shaded boxes show calculated values based on assumed inputs. Capitalizing the cost, to account for the cost of capital during this period of over 13 years, yields a 'capitalized' cost of $1.778 million per NME launch. It is important to note that this model does not include investments for exploratory discovery research, post-launch expenses or overheads (that is, salaries for employees not engaged in R&D activities but necessary to support the organization).
Figure 3 | R&D productivity model: parametric sensitivity analysis. This parametric sensitivity analysis is created from an R&D model that calculates the capitalized cost per launch based on assumptions for the model’s parameters (the probability of technical success (p(TS)), cost and cycle time, all by phase). When baseline values for each of the parameters are applied, the model calculates a capitalized cost per launch of US$1.778 million (see Supplementary information S2 (box) for details). This forms the spine of the sensitivity analysis (tornado diagram). The analysis varies each of the parameters individually to a high and a low value (while holding all other parameters constant at their base value) and calculates a capitalized cost per launch based on those new values for that varied parameter. In this analysis, the values of the parameters are varied from 50% lower and 50% higher relative to the baseline value for cost and cycle time and approximately plus and minus 10 percentage points for p(TS). Once cost per launch is calculated for the high and low values of each parameter, the parameters are ordered from highest to lowest based on the relative magnitude of impact on the overall cost per launch, and the swings in cost per launch are plotted on the graph. At the top of the graph are the parameters that have the greatest effect on the cost per launch, with positive effect in blue (for example, reducing cost) and negative effect in red. Parameters shown lower on the graph have a smaller effect on cost per launch.

determinant of overall R&D efficiency. In our baseline model, Phase II p(TS) is 34% (that is, 66% of compounds entering Phase II fail prior to Phase III). If Phase II attrition increases to 75% (a p(TS) of only 25%), then the cost per NME increases to $2.3 billion, or an increase of 29%. Conversely, if Phase II attrition decreases from 66% to 50% (that is, a p(TS) of 50%), then the cost per NME decreases by 25% to $1.33 billion. Similarly, our baseline value of p(TS) for Phase III molecules is 70%; that is, an attrition rate of 30%. If Phase III attrition increases to 40%, then the cost per NME will increase by 16% to $2.07 billion. Conversely, if Phase III attrition can be reduced to 20% (80% p(TS)), then the cost per NME will be reduced by 12% to $1.56 billion (FIG. 3).

Combining the impact of these increases or decreases in Phase II and Phase III attrition illustrates the profound effect of late-stage attrition on R&D efficiency. At the higher end of the Phase II and III attrition rates discussed above, the cost of an NME increases from our baseline case by almost $0.9 billion to $2.7 billion, whereas at the lower end of the above attrition rates for Phase II and III, the cost per NME is reduced to $1.17 billion.

It is clear from our analyses that improving R&D efficiency and productivity will depend strongly on reducing Phase II and III attrition. Unfortunately, industry trends suggest that both Phase II and III attrition are increasing, given both the more unprecedented nature of the drug targets being pursued, as well as heightened scrutiny and concerns about drug safety and the necessity of demonstrating a highly desirable benefit-risk ratio and health outcome for new medicines. However, maintaining sufficient WIP while simultaneously reducing CT and C will also be necessary to improve R&D efficiency. We discuss these aspects first, before considering strategies to reduce attrition in depth.

Work in process (WIP). We have already emphasized the importance of having sufficient WIP at each phase of drug discovery and development, and have suggested that insufficient WIP, especially in discovery and the early phases of clinical development has contributed to the decline in NME approvals. To further illustrate this point and again demonstrate the impact of Phase II and Phase III attrition on Phase I WIP requirements, we have carried out another sensitivity analysis using these three parameters alone. FIG. 4 shows the impact of varying Phase II and III attrition on the number of Phase I entries per year required to launch a single NME annually. If the p(TS) in Phase II and Phase III are 25% and 50% respectively, approximately 16 compounds must enter Phase I
in Phase III (or even Phase IV). Given the C and CT of a single Phase III unit of WIP ($150 million), almost 10 Phase I molecules ($15 million) can be developed for the same cost, ideally through to proof-of-concept (POC; see discussion of p(TS) below). Reducing late-phase attrition through early POC studies (ideally in Phase I) is therefore crucial to implement this partial solution. The resources (C) saved by lowering Phase III attrition, however, must be redirected to fund sufficient discovery and Phase I/II WIP. Most importantly, advancement into Phase III should be pursued only for those compounds with established efficacy (ideally POC in Phase I and confirmed in Phase II) and a well-defined margin of safety. Ideally, attrition in Phase III should be due primarily to the emergence of relatively rare and unforeseen adverse events. Thus, the key is to have sufficient WIP in the early phases of clinical development to effectively triage and select molecules that will have a higher p(TS) in later-stage development.

The question of how to affordably increase WIP, p(TS) and V without substantially increasing C or increasing CT due to capacity constraints and lack of focus is in our view paramount to improve R&D productivity. This could be accomplished by transforming the R&D enterprise from one that is predominantly ‘owned’, operated and fully controlled by a given company (Fully Integrated Pharmaceutical Company or FIPCo) to one that is largely networked, partnered and leveraged (Fully Integrated Pharmaceutical Network or FIPNet). Traditionally, large pharmaceutical companies have pursued the discovery, development, manufacture and commercialization of their medicines largely by owning and controlling each component. In part, past reliance on the FIPCo model was as much a necessity as a choice. Today, however, the opportunity to partner virtually all elements of R&D through a coordinated and global network or FIPNet could (if effectively managed) substantially improve R&D productivity by affordably enhancing the pipeline from early discovery through to launch. A FIPNet will theoretically allow greater access to intellectual property, molecules, capabilities, capital, knowledge and, of course, talent24–26. Thus, operated as a FIPNet, a given R&D organization will be able to ‘play bigger than its size’ and better leverage its resources to increase WIP across the pipeline, as long as the costs of managing the network do not become prohibitive. Although a full discussion of such a R&D FIPNet is beyond the scope of this article, we would emphasize that it includes many types of partnerships ranging from function-based outsourcing (for example, toxicology or clinical development), to lower C and reduce CT in many cases, to molecule-based risk-sharing partnerships and even partial ownership or equity investments in smaller companies or joint ventures with larger companies. A successfully operated R&D FIPNet should aim to better leverage and proportionally decrease C to affordably increase pipeline WIP, while simultaneously mitigating financial and technical risks.

Much of the discussion on WIP so far has focused on ensuring that there is sufficient WIP to increase NME approvals while addressing the inter-relatedness of the
key productivity parameters. However, all R&D WIP represents expense, albeit an expense that is necessary to generate innovative medicines. Publicly traded companies have a limit to how much they can invest in R&D as a percent of sales and thus the volume of WIP that is affordable. A common mistake is to focus on increasing WIP without regarding V and p(TS). This can easily happen when metrics and reward systems focus solely on the amount of WIP. Ultimately, more WIP without increases in the V or p(TS) associated with the WIP will ensure more output, but increases in R&D expenses will grow proportionately so the strategy will not result in greater productivity. If V and/or p(TS) could be increased, more output or value would actually be derived from less WIP (and expense). At Eli Lilly, we have chosen to focus our view of WIP by adjusting our models to include dimensions of V and p(TS), ensuring we are measuring the value of WIP, not just the amount.

**Value (V).** The effectiveness of drug discovery and development must be increasingly quantified by establishing at the time of product approval a highly desirable health outcome or economic benefit that can be objectively measured in various ways (for example, decreased mortality, morbidity and reduced hospitalizations). Thus, to increase R&D effectiveness it is important to fully understand the ultimate value of a project very early in development and know how this information can be leveraged in individual clinical plans and trade-offs in portfolio decision making.

The determinants of overall value are likely to be different depending on the perspective represented. Patients, caregivers, treating physicians and payers may apply different criteria to determine the value of new drug therapies. For instance, payers are increasingly interested in data from clinical trials with active comparators to establish the benefit of a new medicine versus other therapies (especially generic drugs), whereas regulators primarily depend on placebo-controlled studies to establish efficacy and safety. To optimize the value of a portfolio, it is important that optimal development plans and strategies for each project are formulated early and well before pivotal registration trials are initiated.

We assume that if a given drug treatment provides a higher benefit-to-risk ratio, the potential value of that treatment will be higher\(^2\). Maximizing patient benefit-to-risk ratio and thus potential value can be challenging in many diseases due to substantial clinical and biological heterogeneity. To increase the benefit (and reduce risks) of a drug treatment, it is often important to aim to personalize or tailor the use of the therapy\(^3\). An important aid in the pursuit of tailored therapeutics is the identification of biomarkers that can be used to diagnose the disease and/or identify treatment responders, ideally before or at least after the drug has been given. Biomarkers can also be used to select the right patients, right dose and right duration of treatment, or to avoid exposing patients at risk of a serious adverse event. The use of biomarkers to appropriately select patient subpopulations will also have a positive effect on p(TS) in the later stages of clinical development.

To optimize biomarker development and capture value, it is imperative to ensure biomarkers are developed early and ideally are commercially available at the time of launch. Tailoring therapies to specific patient populations that are predicted to respond on the basis of the presence of a biomarker can be used to reduce development costs C, as illustrated by the development of imatinib (Gleevec; Novartis) or in stratifying patient populations for both clinical development and commercialization, as for trastuzumab (Herceptin; Genentech/Roche). Although many industry pundits have opined that such market segmentation will reduce commercial returns for a given medicine (by reducing market share), so far, the increase in benefit-to-risk (and thus V) has more than offset the reduced market share for many new targeted medicines, especially in oncology.

**Cycle time (CT).** As is evident from our sensitivity analysis (FIG. 3), reducing both Phase III and Phase II CT are also important levers for improving R&D efficiency. Reducing either Phase III or Phase II CT by 50% from our baseline value of 2.5 years to 1.25 years would reduce the C per NME by about $200 million. Similar results were reported by DiMasi\(^4\). Although we think such reductions in Phase II and III CT are unrealistic, even more modest reductions in CT will nonetheless have a significant impact on R&D efficiency. Finally, some aspects of R&D CT, such as the time required for regulatory review (submission to launch in FIG. 3) are less amenable to intervention.

Reductions in R&D CT can be achieved in several ways and have long been a goal in the management of any production system. In predictable production systems, such as product manufacturing, CT reduction principles have been mastered and are routinely applied through techniques such as Six Sigma\(^5\). In less predictable systems, such as pharmaceutical R&D, such approaches to CT reduction cannot be as broadly adapted, but many of the same principles can be readily applied\(^6\).

Considerations for reducing the CT of each phase of discovery and development are often project- and phase-specific. Each phase of development consists of a collection of ‘unit processes’, or individual tasks that consume time and resources. The arrangement of these tasks in time and sequence can often be unique to the individual research project owing to the significant variability introduced by the state of the science associated with the R&D project in question. Because of this inherent variability, CT reduction at a macro level is best achieved through task- and project-specific interventions aimed at reducing non-value-added tasks and ‘wait times’ associated with completing the value-added tasks.

We have identified four key approaches for reducing project-specific C Ts. The first is portfolio selection. When selecting a portfolio of R&D projects, the inclusion of overall project CT as an element of an integrated strategy of portfolio selection should aid in reducing overall C Ts. For example, the CT of pivotal clinical trials are often a function of the disease state or indication being pursued. In some cases, demonstrating the efficacy of a

---

**Imatinib and trastuzumab**
Imatinib blocks the activity of BCR–ABL, a deregulated tyrosine kinase that results from a chromosomal translocation in patients with chronic myelogenous leukaemia, and trastuzumab blocks the activity of HER2/ neu, a receptor tyrosine kinase that is often overexpressed in patients with breast cancer. Patients that are most likely to benefit from each drug can be readily identified before initiating treatment on the basis of the associated biomarkers, which has been invaluable in the development of both drugs and in guiding their use.

**Six Sigma**
A quality management tool that is used to improve the quality of manufacturing and business processes by first identifying and removing the causes of errors or defects, as well as by minimizing variability.
new drug may require many months or years of continued treatment. Moreover, even for drugs for which efficacy can be established relatively quickly (for example, in diabetes), establishing safety, especially cardiovascular safety, may require many months or years of treatment. Thus, considering disease-specific CTSs when selecting molecules or projects for clinical development is crucial for determining the overall movement or flow of a given pipeline.

The second approach to reducing CT is the identification and optimization of the ‘critical chain’ of project tasks. This approach is based on the concepts put forth in the “Theory of Constraints” by Goldratt\(^\text{1}\), wherein the critical chain of tasks is identified and then the sequence, priority and extent of parallel execution is determined prior to project or clinical trial execution. This approach to project management carefully monitors the critical chain and as delays are encountered and incurred, the chain can be reconfigured (in real time) with the aim of recovering any overall project delay. In our experience, when this approach is leveraged against an entire R&D portfolio, it can result in markedly improved CTSs. Thus, CT reduction for the unit work processes that allow for the completion of critical chain tasks is essential.

Once a critical chain of tasks is established, there is the opportunity to improve the business and research processes used to deliver the tasks within the project plan. Each task relies on a process, which is often repeatable and therefore suitable for optimization. Process improvement techniques such as Six Sigma can often be applied to achieve process CT reduction. Examples of this include the introduction of automation for the rapid processing of samples, the use of information technology to support real-time data capture or optimization of clinical trial enrolment. However, elimination of non-value-added activities from the critical chain of tasks is also essential. Examination of the R&D process can result in the identification of activities and tasks that add little or minimal value to desired outputs or outcomes. Often, these tasks are left over from historical approaches to R&D that have become obsolete, or are inefficient bureaucratic procedures associated with the management of R&D that have become institutionalized in an organization, such as multiple serial project reviews by the leadership of each R&D function. When such non-value-added tasks reside on the critical chain of the project, they will delay the project from achieving its milestones.

Finally, the use of adaptive and seamless Phase II and III study designs have proven extremely useful in reducing clinical development CT, generally by reducing or even eliminating non-value-added wait times between phases of development\(^\text{15}\). Selecting those molecules and disease indications for which adaptive and seamless late-stage clinical development is possible can markedly decrease CT and improve \( P \).

As can be seen from our analysis (FIG. 3), not all CT reductions are equal in their ability to reduce overall cost per NME. When applying the approaches above, effort should be prioritized according to both the impact and their feasibility of achieving the proposed CT reduction.

In addition to reducing the overall cost per NME, CT reductions have a positive effect on \( V \) associated with a molecule by allowing greater time on the market under a protected patent.

**Cost (C).** Like CT, an understanding of the C of R&D activities can lead to interventions that can reduce the overall operating expenses required to deliver a successful molecule to the market. The C of an R&D programme can be split into three categories: direct spend on value-added tasks, direct spend on non-value-added tasks and overheads. Focusing on the cost of individual activities that achieve the objectives of an R&D project can yield important opportunities to reduce R&D expense. As R&D involves both the production of knowledge and materials, one must consider how the expense associated with an individual task achieves both production of the requisite information and of the materials that are crucial for downstream activities. Leveraging new technology (such as software tools or laboratory automation) can lead to reduced costs, but care must be taken to eliminate the more tedious and costly methods while introducing new approaches, otherwise costs can become cumulative. As in the case of CT reduction, process improvement methods such as Six Sigma can yield cost savings as well. Finally, accessing lower-cost sources of labour (through outsourcing for example) can yield savings, assuming no compromise of CT or quality.

Identifying and eliminating non-value-added tasks, especially non-molecule C, from R&D programmes can also yield cost-saving opportunities just as it can improve CT. Furthermore, any sufficiently large R&D organization will carry the costs associated with employing and coordinating its employee base. As these overhead costs are spread over the entire portfolio of R&D projects, they have an incremental but often consequential effect on the C of R&D. Because many of the activities that are classified as overheads are indeed essential to the successful operation of a business, they cannot be eliminated entirely. However, in our experience, these activities can be rationalized and reduced to carry the least amount of C considered necessary. Such overhead costs are typically more prevalent in larger, more mature organizations. Unit cost reductions, like CT reductions, can be leveraged to improve productivity (FIG. 3). However, when cutting costs, it is important to be aware not to cut corners that might reduce the \( V \) or \( p(TS) \) of the portfolio.

**Probability of technical success (\( p(TS) \)).** There is little doubt that reducing the attrition rate of drug candidates in clinical development represents the greatest challenge and opportunity for pharmaceutical R&D, and arguably for sustaining the viability of the entire industry. As our sensitivity analyses show (FIG. 3), reducing Phase II and III attrition are the strongest levers for improving R&D efficiency and reducing the costs per NME.

Any given molecule fails because of either technical or non-technical reasons. Non-technical attrition occurs for various reasons, such as a change of strategy or termination of research in a therapeutic area for scientific or commercial reasons, and will not be discussed further, except
to say that it can be a considerable (and sometimes sustained) source of C and lowered R&D efficiency. Technical attrition, on the other hand, occurs when a molecule fails to meet some important success factor, such as failing to demonstrate the requisite safety margin in Phase I or II or expected efficacy in Phase II or III.

Compounds fail for many reasons, but some are more avoidable than others. Poor oral bioavailability, pharmacokinetic properties or toxicity issues that are not predicted by animal pharmacology models or by preclinical ADMET (absorption, distribution, metabolism, excretion and toxicity) studies, resulting in overlap of efficacious and toxic doses (and thus lower than desired margins of safety) are often reasons for Phase I and Phase II attrition.

Given their highly specific target-binding characteristics, fully human or humanized monoclonal antibodies have greatly reduced ‘off-target’ toxicity compared with small molecules, which could help reduce attrition. Any cost-effective approach to predict toxicological liability or undesirable ADME properties preclinically (especially for small molecules) will reduce downstream attrition and several approaches to achieve this are being pursued. In fact, the number (and percentage) of drug candidates successfully reaching Phase I has increased recently, in part due to better preclinical characterization with improved ADMET properties.

Indeed, in their analysis of attrition rates based on data from large pharmaceutical companies between 1991–2000, Kola and Landis highlighted that poor pharmacokinetic properties or bioavailability had become only a minor cause of overall attrition (≤ 10–20%) by 2000, whereas lack of efficacy and low margins of safety were the major causes of Phase II and III attrition. Assuming that approaches for filtering out compounds with inappropriate drug-like properties have advanced since the 1990s, it seems likely that the recent increase in later-stage (especially Phase II) attrition is primarily due to the unprecedented nature of the drug targets (that is, the biological mechanisms) being pursued as well as the increasing safety hurdles (greater benefit-to-risk ratio) required for approval in most parts of the world.

However, as discussed above, Phase II and Phase III attrition rates remain unacceptably high: 66% and 30%, respectively, based on the most recent PBF benchmarking estimates (Fig. 2). Phase II attrition rates in particular have not improved substantially since those reported for the 1991–2000 period (62% attrition rate in Phase II and 45% in Phase III, with some therapeutic areas exhibiting even higher attrition rates). As highlighted by Kola and Landis, clinical attrition rates during the 1990s were higher for central nervous system (CNS) disorders and oncology, with more than 70% of compounds in oncology failing in Phase II and 59% failing in Phase III. The higher failure rates in these areas are in part due to the relatively unprecedented nature of the drug targets being pursued and to the lack of animal models with a strong capacity to predict human efficacy.

There are two key approaches to reduce Phase II and III attrition, which ideally need to be carried out in tandem. The first is better target selection (selection of more validated and druggable targets). The second is the routine pursuit of early POC studies in the clinic, especially in Phase I, for which biomarkers and surrogate endpoints can often be employed. Although these two approaches to reducing attrition and increasing productivity have been highlighted by others, we underscore their importance, and discuss how they can best be achieved below.

First, target selection is a key early step in the drug discovery and development process, as it generally occurs 10–15 years before the launch of a new drug and initiates the commitment of substantial time and resources to determine whether the target and the approach to modulating it are viable. Obviously, target identification and validation leading up to target selection are also important factors, and can confer a competitive advantage to an R&D organization. It was widely predicted that advances in genomics and proteomics, including those resulting directly from the sequencing of the human genome, would yield an abundance of drug targets. However, although many new potential drug targets have been identified by these approaches, far too few have been fully, or even sufficiently, validated so far to have had much impact on pharmaceutical R&D. Nevertheless, we believe that validated targets for drug discovery are now materializing rapidly.

Target selection is also related to the second key approach to reducing Phase II and III attrition: establishing early POC. This requires first choosing a drug target and disease state for which establishing POC early (preferably in Phase I) is feasible. The use of biomarkers or surrogate endpoints (if not clinical endpoints) is often essential for such POC studies. These biomarkers or surrogates can be as simple as assuring target engagement (for example, CNS receptor occupancy by positron emission tomography scanning for neuropsychiatric drugs) or a desired clinical endpoint (for example, the lowering of blood glucose for diabetes or low density lipoprotein (LDL)–cholesterol for dyslipidaemia). A number of clinically relevant endpoints can be ascertained in Phase I (especially in the multidose safety studies or Phase Ib studies), including analgesia, lipid or glucose lowering and weight loss, but developing additional biomarkers of both efficacy and safety for a variety of diseases will be necessary to make early ‘go/no-go’ decisions; such biomarkers are especially needed in oncology.

To illustrate the importance of target selection, coupled with early POC clinical trials, we discuss two recently validated targets that in our opinion will probably result in very effective and commercially important drugs. The first target is the secreted circulating protease proprotein convertase subtilisin/kexin type 9 (PCSK9). Circulating PCSK9 downregulates hepatic LDL receptors and thus increases serum LDL–cholesterol levels. Gain-of-function missense mutations in the PCSK9 gene result in autosomal dominant hypercholesterolaemia and premature coronary heart disease (CHD). Conversely, loss-of-function mutations are associated with very low serum LDL–cholesterol and a striking reduction in the incidence of CHD. Moreover, HMG-CoA reductase inhibitors (statins), which lower LDL–cholesterol, simultaneously
increase PCSK9 levels, reducing their potential efficacy in further lowering LDL–cholesterol. Thus, inhibiting PCSK9 function could dramatically (and safely) reduce LDL–cholesterol as well as augment the beneficial effects of statins. PCSK9 inhibition can conceivably be accomplished in several ways, but perhaps the most direct is with a neutralizing monoclonal antibody, which as noted earlier, could be developed relatively rapidly and with limited off-target toxicity. Such an antibody has recently been reported by Chan and colleagues to dramatically lower LDL–cholesterol in both mice and non-human primates. It should now be relatively straightforward to create a humanized (or fully human) neutralizing monoclonal antibody against PCSK9 for human testing. Importantly, POC with this antibody can be achieved rapidly in Phase I studies by simply measuring LDL–cholesterol, perhaps as soon as following administration of a single dose. Recent clinical and epidemiological studies strongly suggest that lowering LDL–cholesterol below currently accepted targets (with or without a statin) may further reduce the incidence of CHD. It is quite possible that a PCSK9 antibody or some other PCSK9 antagonist could dramatically (and safely) lower LDL–cholesterol and the risk of CHD. However, the key point here is that the validation of this target by human genetics and preclinical studies using a potential therapeutic agent in animals, coupled with the ability to establish POC very early in the clinic, will undoubtedly increase the p(TS) of this molecule before it advances into Phase II and III trials.

The second example is the voltage-gated sodium channel type IX alpha subunit (also known as Na, 1.7 or SCN9A). Like PCSK9, both gain-of-function and loss-of-function mutations in Na, 1.7 have been identified. These mutations result in either increased sensitivity to painful stimuli (paroxysmal extreme pain disorder and primary erythermalgia) or channelopathy-associated insensitivity to pain, respectively. The latter is a rare autosomal recessive disorder in which affected individuals are insensitive to physical pain from birth. Importantly, these individuals have otherwise normal sensory perception, although the consequences of complete and persistent insensitivity to pain can be quite detrimental. Nonetheless, a selective and reversible Na, 1.7 antagonist, perhaps one that inhibits channel activity in a state-dependent manner, will probably prove to be a very effective topical or oral analgesic agent. Non-selective sodium channel blockers, such as the local anesthetic lidocaine and certain anticonvulsants such as phenytoin and carbamezepine, may in part work through Na, 1.7, but finding more selective antagonists could revolutionize the treatment of pain. Moreover, as was the case for PCSK9, clinical POC can be rapidly obtained in Phase I using various human pain models. These two drug targets have one important feature in common. Mutations in their genes cause a loss of function that results in a clinical phenotype (low LDL–cholesterol or insensitivity to pain) that could be mimicked by a drug with potentially desirable therapeutic consequences. The validation of these two targets derives primarily from human genetics, coupled with a very good (but admittedly still incomplete) understanding of their biological mechanisms and clinical consequences. We estimate that there are now at least a dozen or more drug targets with similar human validation and this number will undoubtedly grow as advances in functional genomics continue to materialize.

Using these two approaches (better target validation and early POC studies) we have estimated that the p(TS) of Phase II compounds can be increased to approximately 50%. As can be seen in our sensitivity analysis (Fig. 3), reducing Phase II attrition by this much will by itself lower the cost per NME by approximately 50%.

In our view, resolution of technical uncertainty early in development, especially whether or not a molecule engages its target and has desired pharmacological activity in humans is necessary to improve R&D productivity. We refer to this as our ‘quick win, fast fail’ paradigm of drug development (Fig. 5), whereby the ‘sweet spot’ of R&D resides prior to Phase II with a heavy emphasis on having sufficient discovery capacity and capability to assure selection of validated targets, and sufficient discovery WIP to yield enough good Phase-I-ready molecules. A resolute focus and portfolio metric is to establish POC in the clinic very early, preferably in Phase I. This can reliably be achieved for most, but obviously not all, drug candidates.

We have implemented an alternative clinical development model called Chorus, which has served as a highly cost-effective means for establishing POC and has, in part, helped to increase the estimated p(TS) of...
**Summary and additional considerations**

Without a substantial increase in R&D productivity, the pharmaceutical industry’s survival (let alone its continued growth prospects), at least in its current form, is in great jeopardy. Much has been written on this subject and multiple contributing factors and potential solutions have been proposed. Our econometric model of the current cost of discovering and successfully developing a single new medicine reveals the major contributors to the escalating costs and the declining number of launches. Importantly, our parametric analyses further reveal where the greatest improvements in productivity must occur.

We also offer a number of potential solutions, which if effectively implemented could dramatically reduce costs and importantly increase the flow of truly innovative new medicines to patients. For example, using our model parameters (Fig. 2) and sensitivity analyses (Fig. 3), we can elaborate a series of interventions to reduce the cost per NME by 28% and 50% respectively. Fig. 6 illustrates the potential interventions designed to increase productivity by primarily reducing CT and C, as well as by shifting attrition to early stages of clinical development, increasing the overall p(TS) in Phase II and III. Although the more extreme case of a 50% reduction in cost per NME is ambitious, it is in our view not unrealistic and may be necessary to sustain a viable R&D business model.

As the pharmaceutical industry transitions from an era of ‘me-too’ or ‘slightly me-better’ drugs to one of highly innovative medicines that result in markedly improved health outcomes, it must, as our model dictates, re-focus its resources (money and talent) on discovery research and early translational medicine. Although the scientific substrate for drug discovery has never been more abundant, a more complete understanding of human (disease) biology will still be required before many true breakthrough medicines emerge.

We also emphasize that there are other important elements in a highly productive R&D enterprise that are not included in our model, most of which are necessary, although not sufficient to assure success. As underscored by Cuatrecasas, fostering scientific creativity and being opportunistic for serendipitous scientific and medical findings are clearly important elements of past and future success in the pharmaceutical industry. Over-managing or reducing CT and C, and particularly p(TS) by 17% in Phase I and by 47% in Phase II by 47%.

An even larger 50% reduction in cost per NME could be achieved by additionally reducing CT by 20% across Phases II and III, a further reduction in cost by 10% across all phases, and an increase in Phase III p(TS) by 14% without negatively affecting other parameters.

However, one risk of using this quick win, fast fail paradigm is that false-negative POC data (type 2 errors) could offset some or all of the cost savings generated by shifting attrition earlier. Thus, it is important to ensure that the additional studies that form the basis for the early decision have a relatively low uncertainty, which is not possible for all projects.

Finally, we consider that one of the major contributors to high Phase III attrition is simply premature advancement of NMEs into Phase III; that is, prior to establishing efficacy and an acceptable safety profile in Phase I (POC) and Phase II. The latter is often the result of poor discipline in portfolio management combined with perceived near-term business imperatives.

**Figure 6** | **Productivity interventions yielding a substantial reduction in the cost per new molecular entity.** This analysis shows the potential beneficial effect on capitalized cost per new molecular entity (NME) launch resulting from improvements in a number of R&D parameters (cost, cycle time (CT), probability of technical success (p(TS))) as calculated using the R&D model described. A 28% reduction in cost per NME could be achieved through the aggregate actions resulting in a 9% reduction in CT from lead optimization to Phase I, a 10%, 15% and 20% reduction in cost for clinical development Phases I, II and III respectively, and a shift in attrition from Phase II to Phase I (decrease p(TS) in Phase I by 17% and increase p(TS) in Phase II by 47%). An even larger 50% reduction in cost per NME could be achieved by additionally reducing CT by 20% across Phases II and III, a further reduction in cost by 10% across all phases, and an increase in Phase III p(TS) by 14% without negatively affecting other parameters.
even micro-managing R&D by those with little scientific–medical expertise or experience in R&D may well have contributed to recent pipeline gaps and reduced R&D productivity. Moreover, fostering such scientific creativity and medical opportunism is often challenging, and sometimes incompatible with the short-term and often aspirational goals of business-driven enterprises. Most importantly, in our opinion, there is no substitute for scientific and clinical intuition for any successful R&D organization, whether it be in the initial selection of drug targets, the design of a clinical trial or the interpretation of clinical data in order to advance a molecule into late-phase clinical development. Simply stated, good ‘process’ will never be a substitute for good people or good science. However, we firmly believe that both good process and good science are not only compatible, but together will yield the greatest return on R&D investments and thus, have the greatest impact on R&D productivity.

4. A candid assessment by the former CEO of GlaxoSmithKline of the downsides of the industry’s process culture.
12. An insightful retrospective on the biotech industry, and the reasons why it has not yet lived up to its promises.
14. A contemporary assessment of the costs to discover, develop and launch new molecular entities.
21. Booth and Zemmil framed the issue of productivity in the pharmaceutical industry and provide a qualitative discussion of some of the drivers and potential solutions.
24. This article provides an in-depth analysis of attrition by phase and therapy area, and offers a scientific approach to reduce Phase II and Phase III attrition.
26. A seminal paper on the hurdles that must be overcome to reverse the decline in R&D productivity.