

# Invention reinvented

McKinsey perspectives on pharmaceutical R&D



# Invention reinvented

McKinsey perspectives on pharmaceutical R&D  
2010

Head of Global and North American Pharmaceutical and Medical Products R&D Practice  
Rodney Zimmel

Head of European Pharmaceutical and Medical Products R&D Practice  
Mubasher Sheikh

For more information please contact  
[pharma\\_r&d@mckinsey.com](mailto:pharma_r&d@mckinsey.com)

**London**

No. 1 Jermyn Street  
London  
SW1Y 4UH  
United Kingdom  
Tel: +44 (20) 7839 8040

**New Jersey**

600 Campus Drive  
Florham Park  
NJ-07932-1046  
United States  
Tel: +1 (973) 549 6600

**New York**

55 East 52nd Street  
21st Floor  
New York, NY-10022  
United States  
Tel: +1 (212) 446 7000

**Silicon Valley**

3075A Hansen Way  
Palo Alto  
CA-94304  
United States  
Tel: +1 (650) 494 6262

**Zurich**

Hotelstraße  
Zürich-Flughafen  
CH-8060  
Switzerland  
Tel: +41 (44) 876 8000

# Contents

1. **The road to positive R&D returns**  
*Scientific innovation is not the only way to higher R&D productivity. Attention to the familiar management areas of cost, speed, and decision making can still reap rewards.*  
Eric David, Tony Tramontin, and Rodney Zimmel
  
7. **The secret of high productivity in the research lab**  
*What drives research productivity? An understanding of how the world's most successful laboratories operate reveals some answers.*  
Mark Beards, Michael Edwards, and Mubasher Sheikh
  
15. **The future of drug development: advancing clinical trial design**  
*Traditionally, drug development has been conducted in distinct—and increasingly expensive—phases. A more integrated model that uses adaptive design approaches to enhance flexibility and maximize the use of accumulated knowledge could improve the quality and reduce the cost.*  
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 Evers, Edd Fleming, Navjot Singh, Tony Tramontin, and Howard Golub
  
33. **Using IT to speed up clinical trials**  
*The first wave of IT improvements solved some problems, but others remain. Now it is time for a more comprehensive approach.*  
Sam Marwaha, Samir Patil, and Navjot Singh
  
41. **Quality gates: raising the standard of pharmaceutical technical development**  
*Good technical development can improve quality, lower costs, and get products to market more quickly, yet few pharmaceutical companies manage it well. Quality gates can transform the process.*  
Michele Holcomb, Martin Møller, and Paul Rutten
  
47. **Market-access-minded development**  
*A new way of conceiving clinical development will uncover the drugs bound for commercial success early on. Think value.*  
Matthias Evers
  
51. **Changing with the times: how pharmaceutical companies should respond to regulatory change**  
*Drug companies need to make organizational changes to meet the latest patient safety regulations in the United States. But as the regulations evolve, companies will also need to work openly with all stakeholders to ensure the right outcome.*  
Maria Gordian and Richa Pande
  
57. **The anatomy of attrition**  
*Research tracking the attrition rates of more than 3,000 compounds in each phase of development reveals recent drug failure trends.*  
Navjot Singh, Rodney Zimmel, Maria Gordian, and Áine Denari
  
65. **What gives drugs the X factor?**  
*Despite the long development process, more than two-thirds of new drugs launched on the market are commercial failures. Recent research uncovers what distinguishes those that succeed.*  
Ramesh Hariharan and Navjot Singh
  
73. **A special kind of success**  
*Mainstream pharmaceutical companies are turning their attention to specialty drugs. Which factors are most important for commercial success or failure?*  
Mark Gudiksen, Edd Fleming, Laura Furstenthal, and Philip Ma
  
81. **R&D in emerging markets: a new approach for a new era**  
*Gone is the time when R&D in developed markets could meet the pharmaceutical needs of emerging markets. Local presence is required.*  
Michael Edwards
  
87. **The microeconomics of personalized medicine**  
*Personalized medicine promises to increase the quality of clinical care and, in some cases, decrease health care costs. The biggest hurdles are economic, not scientific.*  
Jerel Davis, Laura Furstenthal, Amar Desai, Troy Norris, Saumya Sutaria, Edd Fleming, and Philip Ma
  
101. **Authors' contact details**

## Introduction

### Invention reinvented

In the eyes of patients and other stakeholders, the pharmaceutical industry exists to discover new medicines that go on to become standard treatments. The faltering economics of R&D productivity are jeopardizing that mission: R&D expenditure is not delivering.

According to analysis, costs have ballooned to more than \$1.6 billion per new drug approved, compared with \$1 billion in 2005, attrition rates remain stubbornly high, and many of the drugs that do eventually reach the market fail to gain broad acceptance. Today, only 30 percent of drugs launched earn an acceptable rate of return—a rate that is likely to deteriorate further given the increasing demands of payors and access agencies that want more value for their money. In 2008, for example, 96 percent of drugs reviewed by the French authorities were classified as having “minor differentiation at best,” meaning limited market access. And between 1998 and 2008 in the United Kingdom, almost 60 percent of drugs from the top ten pharmaceutical companies were given negative or restricted access. The recent global economic crisis can only make payors even more sensitive to costs.

Many have long called for what is effectively a new R&D paradigm—although apparently to little effect, given that drug approvals continue to fall. Today however, we sense real change as companies learn to adapt to a far harsher environment. Invention is being reinvented.

In this collection of articles, we set out our thoughts on some of the elements of that reinvention. We examine the untapped scope for improving productivity not through scientific innovation, but through better management of R&D and the use of IT in clinical trials. We publish original research on why so many drugs fall out of the development pipeline, and on what makes some laboratories more productive than others. We analyze the factors that account for commercial success, and look at how pharmaceutical companies are changing the way they work and think to meet new regulatory safety requirements and to gain market access. We assert that clinical development can be improved by moving from the traditional sequential approach to a more integrated model. And we investigate emerging opportunities—dissecting the economics of personalized medicine, and explaining why now is the time to refocus R&D on the needs and preferences of developing markets.

We hope you will find these ideas stimulating and welcome your comments. Please send them to [pharma\\_r&d@mckinsey.com](mailto:pharma_r&d@mckinsey.com), or contact any of the authors individually.



Rodney Zimmel  
Director, New York office



Mubasher Sheikh  
Principal, London office

## The road to positive R&D returns

Scientific innovation is not the only way to higher R&D productivity. Attention to the familiar management areas of cost, speed, and decision making can still reap rewards.

Eric David, Tony Tramontin, and Rodney Zemmel

As productivity in the pharmaceuticals industry has fallen, so calls for a “new paradigm” that will radically change pharmaceutical R&D have increased (Exhibit 1). The trend has been to view diminishing returns as a science problem. But while scientific innovation is certainly part of the solution, management should not overlook other, more familiar means of value creation. Increased attention to costs, speed of development, and decision making could increase the internal rate of return (IRR) of an average small molecule from around 7.5 percent—which is less than the industry’s cost of capital—to 13 percent (Exhibit 2).

### Sizing the challenge

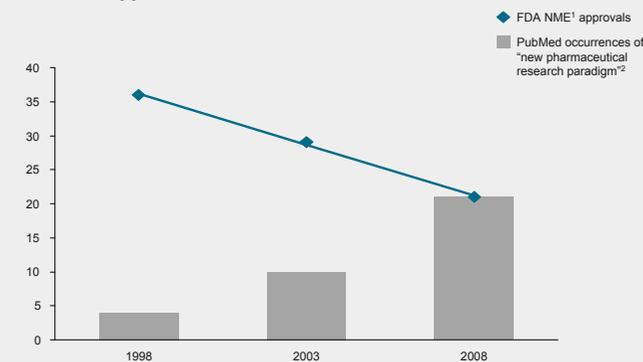
We modeled the estimated average return on R&D investment for a typical small-molecule compound and for a typical biologic. With net present value (NPV) of minus \$65 million and an IRR of 7.5 percent, the model suggested that present-day returns for an average small molecule fall below the cost of capital (Exhibit 3). By contrast, between 1997 and 2001, the return approached 12 percent.

The key factors driving this downward trend in productivity are well known. Industry interviews and analysis of the Pharmaprojects database indicate that, over the past decade, the overall probability of success (POS) for small molecules has fallen by five percentage points, and the time required for R&D has increased by between 12 and 18 months. Furthermore, R&D costs have recently risen by about 8 percent annually, while prices worldwide are under pressure.

It could be argued that companies should shift much of their R&D investment to biologics,

EXHIBIT 1

### More talk, fewer approvals



<sup>1</sup> New molecular entity.

<sup>2</sup> PubMed search for "new pharmaceutical research paradigm." The term "cardiac surgery" was used as a control over the same time period to ensure trends were not simply due to changes in the number of publications available in PubMed over time. Source: NME data from Parexel's *Pharmaceutical R&D Statistical Sourcebook 2008/2009*

as the average biologic currently offers a greater return (NPV of \$1.26 billion, IRR of 13 percent) owing to higher average peak sales and slower decay of sales following loss of exclusivity. But given the limited number of such molecules and the expected erosion of returns as competition from biosimilars mounts, increased investment in biologics alone is not the solution. Rather, more careful management attention to reducing costs and accelerating time-to-launch, and better decision making over when to remove poorly performing drugs from the portfolio and which compounds to invest in, will reap significant rewards.

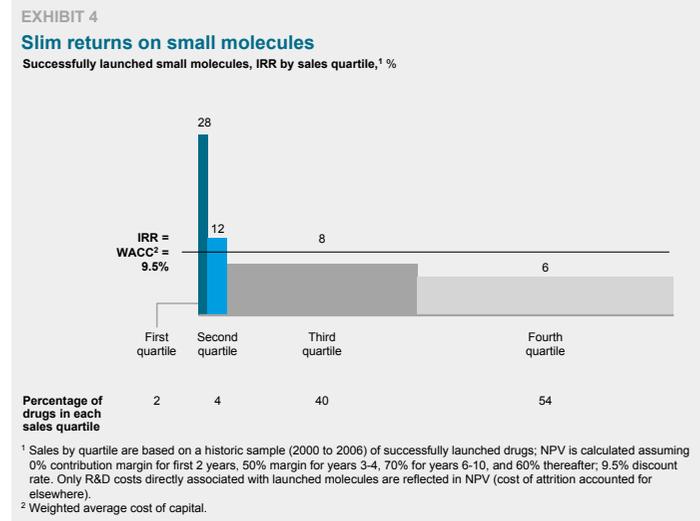
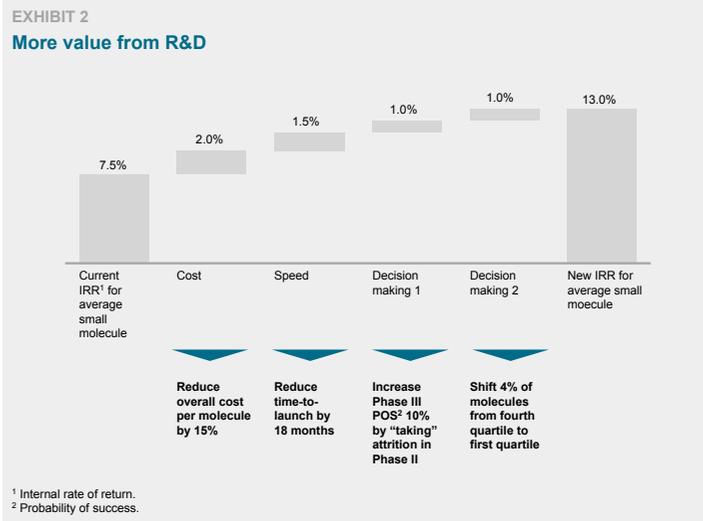
## Levers for change

### Costs

Although most companies have made progress in reducing costs, efforts too often focus on the obvious organizational and procurement issues. Successful companies generally employ broader strategies.

One approach is for companies to change what they are doing, not just how they do it. For example, companies that consistently over-power clinical trials could reduce the number of patients per trial. Our experience also suggests that R&D costs could be reduced by between 5 and 10 percent through more

aggressive outsourcing of selected non-core activities to low-cost geographies. A second approach is to reduce the costs associated with drug failures. Companies generally design R&D programs for success, even though the majority of programs will fail. Costly, two-year carcinogenicity studies, for example, are often initiated before a compound reaches proof of concept at the end of Phase IIa, and this expenditure is wasted if the compound fails (as it is likely to do). Eli



Lilly's Chorus unit represents one effort to reduce costs by focusing on the activities that truly reduce the risk of failure of a compound on the way to proof of concept. The cost of failure can also be lessened by sharing risk with another party, such as another pharmaceutical company, a contract research organization, or investors. Such strategies can together reduce the overall cost of R&D by 15 percent or more, increasing the NPV of average small-molecule projects by about \$250 million and raising the IRR of R&D by some two percentage points.

### Speed

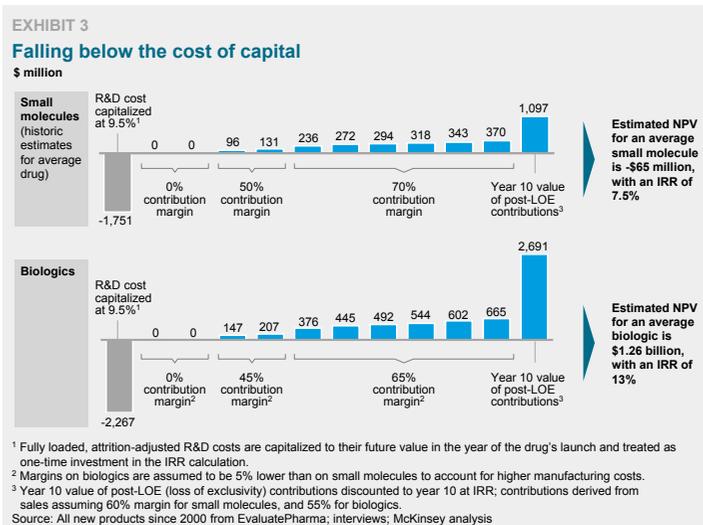
For medicines that make it to market successfully, our modeling indicates that each six-month delay to launch can mean a loss of almost \$100 million in NPV, or a reduction of 0.5 percentage points in IRR. This number is obviously much higher for top-performing drugs. Yet opportunities exist to address inefficiencies such as poor planning of clinical development, slow patient recruitment, and suboptimal site and investigator management. We modeled the effect of accelerating a development

program by a conservative 18 months. This increased the NPV of an average compound by about \$190 million, raising the IRR by 1.5 percentage points. Some companies have done much better: Merck accelerated the launch of the diabetes drug sitagliptin (Januvia) by three to four years by employing novel parallel development techniques. But gains in speed cannot come from short cuts: the key to capturing value from

program acceleration is choosing the right programs to accelerate.

### Decision making

R&D leaders grapple with decisions about program termination, acceleration, resourcing, and prioritization. Project-termination decisions are especially difficult and can cost a company hundreds of millions of dollars if made too late. The current high attrition rate in Phase III trials suggests that companies have overlooked or ignored key signals, and in some cases made poor decisions about aspects over which they have substantial control. Our analysis indicates that of 106 reported Phase III failures from 1990 to 2007, 45 percent were due to insufficient efficacy of a drug versus a placebo, and 24 percent to insufficient differentiation versus standard of care. It is easy to scrutinize decision making with the benefit of hindsight, but R&D leaders can increase returns by identifying and removing poor performers from the portfolio earlier in development. Many organizations still advance compounds for the wrong reasons: because of "numbers-focused"



incentive systems, because they fail to understand how product differentiation is increasingly driving reimbursement, or because they have traveled too far down the development path.

Many companies have started to restructure to address these issues. Eli Lilly's Chorus unit, GlaxoSmithKline's Discovery Performance Units, and Pfizer's smaller, more focused therapeutic areas are a few examples. If these efforts enable R&D leaders to make better decisions and shift compound attrition to earlier stages, the impact will be substantial. "Taking" attrition earlier during Phase II could increase Phase III survival by ten percentage points—comparable to survival rates between 1997 and 2001—and increase IRR by up to one percentage point.

Another key aspect of R&D decision making is the choice of compounds in which to invest. We used data for successfully launched drugs between 2000 and 2006 to divide products into quartiles based on their returns. As shown in Exhibit 4, the top two quartiles of launched molecules account for most of the value creation. A top-quartile small molecule has an IRR of 28 percent compared with the 7.5 percent of an average small molecule. A top-quartile biologic has an IRR of 33 percent compared with 13 percent for an average biologic. Second-quartile molecules have an IRR of 12 percent for a small molecule and 15 percent for a biologic. No company is likely to be able to identify

top-quartile drugs consistently: our analysis shows that, in any given portfolio of small molecules, only 2 percent of drugs will be top-quartile sellers, while 54 percent will be fourth-quartile sellers. However, by shifting even 4 percent of compounds from the fourth quartile to the top quartile, the average IRR would increase by one percentage point.

### Implications for R&D leaders

A consistent, aggressive, and simultaneous focus on costs, speed, and decision making can raise the IRR on an average small molecule from 7.5 percent to about 13 percent. For a typical portfolio of a leading pharmaceutical company, assuming a composition of 75 percent small molecules and 25 percent biologics distributed across various phases of development, this would raise the portfolio return to between 14 and 15 percent, from between 9 and 10 percent currently.

Previous industry experience suggests such goals are attainable: from 1997 to 2001, the return on the portfolio described above would also have been 14 to 15 percent, driven by a higher POS and shorter development times. Although the current environment is tougher, managers are not yet fully exploiting the value-creation levers described here, and moderate improvements can substantially increase returns. An IRR of 14 to 15 percent on R&D might not sound like hitting the jackpot, but over a large portfolio it would create considerable value.

A version of this article, entitled "Pharmaceutical R&D: the road to positive returns," was first published in *Nature Reviews Drug Discovery*, August 2009, Volume 8, pp. 609–10. For further details of the analysis and methodology, please contact the authors.

Eric David ([eric\\_david@mckinsey.com](mailto:eric_david@mckinsey.com)) and Tony Tramontin ([tony\\_tramontin@mckinsey.com](mailto:tony_tramontin@mckinsey.com)) are associate principals in McKinsey's New York office, where Rodney Zimmel ([rodney\\_zimmel@mckinsey.com](mailto:rodney_zimmel@mckinsey.com)) is a director.



## The secret of high productivity in the research lab

What drives research productivity? An understanding of how the world's most successful laboratories operate reveals some answers.

Mark Beards, Michael Edwards, and Mubasher Sheikh

Increasing productivity is one of the fundamental challenges of life sciences research. The work is complex, lengthy, and costly, and the failure rate high. But successful research can lead to disease-beating medicines and significant financial rewards for the research organizations involved. In the past, companies have tried to improve productivity through extensive, top-down initiatives, such as reorganizing research or investing heavily in new technology platforms. McKinsey, through its "SuccessLab" initiative, has now taken a different approach, setting out to understand what drives research productivity from the bottom up—from the laboratory itself.

To grasp how the world's leading research laboratories organize and manage their research, we interviewed 12 world-class academic innovators, including Nobel prize winners, McArthur Genius Award winners, and some of the world's most highly cited scientists.<sup>1</sup> We wanted to understand what might differentiate their approach from that used in other, less productive laboratories, and whether there were key practices and processes that accounted for their success. A clear pattern emerged from the interviews.

We then examined the practices of 15 research laboratories in industry with

different levels of performance, and several more academic laboratories. The aim was to compare these laboratories' practices with those identified in the first phase of the research (see sidebar, "Research participants").

As a result, we identified many similarities in the approaches adopted by all the most productive laboratories, or "top labs." Although the top labs were often conducting quite different kinds of research, elements of their approach in five areas—strategy decisions, talent management, portfolio and project management, problem solving, and collaboration—were remarkably similar. Be they in academia, the pharmaceutical industry, high technology, or industrial chemical manufacturing, top labs organize themselves to ensure they have the right team working with a clear focus on a shared set of activities, and that researchers spend as little time as possible on administration and management. The result is higher productivity.

Although there are structural differences in how the various top labs try to achieve this outcome, the underlying principles are often consistent. Below we describe these principles, which together, our research suggests, amount to a best-practice approach to higher productivity.

## Research participants

### 12 Academic labs

- Dr Shizuo Akira, *Osaka University*
- Prof. Carolyn Bertozzi, *UC Berkeley*
- Prof. Sam Danishefsky, *Sloan-Kettering Cancer Center and Columbia University*
- Prof. Ulrich Hartl, *Max Planck Institute of Biochemistry*
- Sir Tim Hunt, *Cancer Research UK*
- Prof. Steven Jackson, *University of Cambridge*
- Prof. David Liu, *Harvard University*
- Prof. Eiichi Nakamura, *University of Tokyo*
- Prof. Svante Pääbo, *Max Planck Institute for Evolutionary Anthropology*
- Dr Luca Scorrano, *University of Geneva and Venetian Institute of Molecular Medicine*
- Prof. George Whitesides, *Harvard University*
- Sir Greg Winter, *MRC Laboratory of Molecular Biology*

### 12 Pharmaceutical labs, covering:

- Discovery
- Medicinal chemistry
- Technology
- Structural sciences
- Lead evaluation

### 3 Additional labs

- Consumer electronics
- Semi-conductors
- Polymers

The research project is continuing and the database expanding. To date, more than 1,000 scientists have participated.

## Strategic decisions

Clear, three- to five-year strategies, focused on portfolios of interconnected projects, were a shared characteristic of the top labs. Clarity of strategy enables

effective decision making on allocation of resources, the building of new capabilities, and communication of a lab's role among its own staff, throughout the wider organization, and externally. The constant search for ways to create competitive advantage through emerging technologies was another recurring strategic feature.

## Talent management

In evaluating researchers, top labs rate intrinsic intellectual capability, scientific curiosity, and general problem-solving skills higher than specific technical knowledge. There are of course occasions when a specific technical capability is required. But even in these situations, top labs will try to hire people with the ability to adapt to new roles as the focus of the lab evolves. To bring a range of perspectives and approaches to their work they recruit widely, from different countries and institutions and from diverse academic and commercial backgrounds.

One laboratory head believes that “regular interviews are totally useless” as a way of assessing candidates and prefers to rely on recommendations from collaborators and former colleagues or to ask applicants to work in the laboratory for a trial period. Other recruiters supplement interviews and technical assessments with problem-solving sessions, perhaps asking candidates to spend a morning devising answers to a specific question or working in the laboratory with the team. The latter serves the dual purpose of indicating how well the candidate will fit in. With this in mind, some laboratories arrange a social occasion before hiring, and the views of existing team members are always sought before a decision is made. Aware that courtship is a two-way street, top labs seek to attract the best candidates by nurturing their profile

and reputation through working with academic institutions, offering PhD sponsorship and post-doctoral posts, presenting at conferences, taking active roles in industry groups, and publishing. (The fact that they have a clear strategic focus makes reputation building an easier task.) Internally, too, they work to attract talented individuals by way of secondments and transfers, particularly in industry. Internal publications, poster sessions, and meetings serve to bring the laboratory to the attention of the wider organization.

Once they are in place, new recruits are given formal and informal support and advice to help them to adopt the laboratory's high standards and practices and to become useful as quickly as possible. Existing team members are expected to commit significant time to the one-on-one apprenticeship of new joiners (their end-of-year evaluations take account of this), and to assist in the continuous mentoring of junior lab members. Encouraged in this way, new hires and juniors feel their careers are developing and that they are helped to reach their full potential. For a laboratory to do otherwise is to risk losing them.

Financial incentives are transparently linked to certain achievements—an exceptional piece of research, or publication in a leading journal. At the other end of the scale, poor performance is not tolerated for long, and researchers are asked to leave if they do not respond to efforts to improve their attainment.

Despite the efforts they make to build the right teams, top labs are not overly concerned about staff turnover, seeing it as an important way of preventing their problem-solving culture from stagnating. Some even encourage

turnover, although this can be more difficult to achieve in industry than academia. Seconding team members to different parts of the organization and exposing them to the different areas of focus within the laboratory's portfolio can help to keep teams fresh.

## Project and portfolio management

The top labs design their portfolios of projects to be interlinked, so that they are both additive, in that the lab reaps benefits from their intellectual scale, and synergistic, in that each project might uncover insights that prove valuable to another. Competitive advantage arises in both respects.

Project teams are assembled to incorporate the mix of skills needed to address particular problems, and it is these, rather than functional groups, that are the focus of the organization. Individual members too are expected to have expertise in several disciplines—the same person might run biophysical and cell biological assays on p53, for example—rather than depth of functional skills in one area, enabling the team to work flexibly to address its needs at any point. Subject to the laboratory's requirements, researchers are allowed a degree of freedom to choose which projects they work on, and staffing allocation is discussed with all researchers before decisions are made.

Once a team is assembled, members set about their project with a clear plan that details what should be achieved by when and the mechanism for reviewing progress. The plan is used to forecast when particular resources or equipment might be required and to anticipate potential bottlenecks or

delays. Simple visual tools are used to share information in a way that is quickly understood: for example, what is happening and when, or any likely clashes in demand for equipment.

But if all does not go according to plan, if milestones are missed, top labs do not shy away from tough decisions. Although a plan might sometimes warrant modification, they are keen to terminate sooner rather than later any project that is not showing results. Pouring more resources into a project that is unlikely to add value, or allowing it to continue at a lower level—because researchers are reluctant to let go or because they feel they might have missed something—robs other projects that have greater potential.

### Problem solving

The best laboratories define a project by the specific solution sought, then use a variety of approaches to solve the problems along the way. They are careful to build on the platforms and techniques in which they have proven strengths in order to see where they can extend into new areas. In this way, their areas of expertise evolve.

Before running experiments, they ensure they have an initial hypothesis. To avoid mission creep, they may even write the first draft of the final paper abstract at the outset—even if they do not intend to publish the work externally—and redraft the paper as work progresses. One academic laboratory uses the draft final paper as the starting point for all reviews.

Notwithstanding these controls, laboratory heads are keen to give teams a degree of autonomy and not over-define the specific scientific approaches they should use, as doing so risks demotivating researchers and smothering potentially

innovative ideas. Instead, they work with the broad group to outline the general approach. Their role is to ensure that the projects continue to contribute to the overall goal of the laboratory and are synergistic, not to micro manage unless absolutely necessary. Laboratory heads also encourage their researchers to spend time on projects driven by their own personal interest. This could be defined as a percentage of their time or as a set period each week: one laboratory sets aside Friday afternoons for work on personal initiatives. Allowing researchers this freedom helps maintain their passion for work, but can also ensure innovative ideas are not overlooked. Some laboratories have seen their Friday afternoon projects develop into major research initiatives.

Finally, the laboratories we examined have a healthy attitude to failure, seeing failed experiments as excellent learning opportunities. They take time to review raw data thoroughly, and might invite the head of another laboratory or team to do the same to help tease out any useful conclusions.

### Collaboration

An open, sharing culture is important for productivity. Lab protocols and manuals are explicitly codified and shared, but top labs also call regular meetings for teams to share knowledge and challenge it, and to discuss the difficulties they face. They also create opportunities for teams to meet informally, perhaps over coffee or lunch, seeing this as an important way to identify how different teams can support one another.

One academic laboratory holds a Journal Abstract Club, at which all abstracts from the top 30 scientific journals are

screened and six chosen to be presented at the next club meeting. Each club member presents once every two to three weeks. The laboratory head says he sees the club as a way of keeping everyone informed about what's going on in the wider world, as well as better informed about what not to do.

Generally, external collaboration with other academic and industry laboratories is not wide spread, and might be instigated only to address a specific problem. The reason is the perceived need to protect a laboratory's innovations. However, external collaboration can be a valuable source of ideas given the correct agreement structure, and top labs see it as an important element of their work that can enable a wider group of researchers to be brought to bear on the biggest challenges. These collaborations are quite different to the transactional collaborations used to provide a specific, routine service: they tend to be long-term relationships whereby researchers work in one another's laboratories and meet regularly to share ideas and solve problems. These collaborations are often with groups headed by laboratory alumni, particularly in the academic sphere.

Top labs are also aware of how physical proximity can promote collaboration, organizing departments so that different teams and disciplines work closely together and even sit in the same areas. In one laboratory, chemists and biologists share a space for their desk work, with the chemistry lab on one side and the biology lab on the other. Large coffee areas and a lunch room for the laboratory staff encourage the culture of sharing and networking between teams. "Hospitals and labs are the opposite: at hospitals you want to minimize contact to prevent diseases spreading; in labs

you want things to spread as much as possible like infectious diseases. You need crowdedness," is the way one head of an academic laboratory puts it.

### Differences between laboratories

Understanding what top labs do to succeed makes it relatively easy to see why others that fall short of best practice are less productive. If, for example, a laboratory has only a short-term strategy, it is likely to struggle to identify where to build capabilities, to link projects to get the benefits of shared learning, and to articulate what the laboratory should be known for. If it fails to recruit researchers according to best-practice criteria, or if it moves poor performers to other projects rather than dismissing them, it might limit the range of its approaches to problem solving and undermine performance. And if it does not link its portfolio of work, it also fails to make use of the intellectual scale of the organization and misses opportunities to bring new insights to solving problems. It might explain away its decision not to invest in "side projects" by invoking budgetary and business pressures, but the outcome remains the same: less enthusiastic researchers and missed potential. Budgetary concerns may also affect the extent to which average labs can design their facilities to encourage internal collaboration: teams working in older facilities are often separated from other parts of their laboratory group, and might even be located on different sites. But again, whatever the cause, the result is likely to be lower productivity because of lost opportunities to share learning and solve problems together.

We are not suggesting that one size fits all. Academic and industry laboratories have different requirements, and our research

did not identify a single laboratory that followed all the best practices described here. Nevertheless, the research identifies the common approaches the world's

top labs take to make their research activities productive, from the bottom up. It should prove a useful resource to those wishing to emulate their performance.

For more information about the research, analysis, and its application, please contact [successlab@mckinsey.com](mailto:successlab@mckinsey.com)

**Mark Beards** ([mark\\_beards@mckinsey.com](mailto:mark_beards@mckinsey.com)) is a consultant in McKinsey's London office, where **Michael Edwards** ([michael\\_edwards@mckinsey.com](mailto:michael_edwards@mckinsey.com)) is an associate principal, and **Mubasher Sheikh** ([mubasher\\_a\\_sheikh@mckinsey.com](mailto:mubasher_a_sheikh@mckinsey.com)) is a principal.

---

<sup>1</sup> We identified 30 internationally prestigious awards in medicine, biology, and chemistry, then identified the recipients of these awards between 2000 and 2008. Their innovation productivity was measured by their impact-adjusted publication productivity (Hirsh index divided by years of publishing activity) and IP-generation productivity (US patents divided by years of publication activity). Twelve of the top-quartile scientists agreed to take part in the research.

## The future of drug development: advancing clinical trial design

Traditionally, drug development has been conducted in distinct—and increasingly expensive—phases. A more integrated model that uses adaptive design approaches to enhance flexibility and maximize the use of accumulated knowledge could improve the quality and reduce the cost.

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 Evers, Edd Fleming, Navjot Singh, Tony Tramontin, and Howard Golub

Pharmaceutical innovation is increasingly risky, costly, and at times inefficient—with obvious implications for productivity.<sup>1-3</sup> Estimates of the average cost of bringing a new drug to market range from \$800 million to \$2 billion, in part owing to late-stage failures and the rising costs of Phase II and III trials.<sup>4-9</sup> 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,<sup>10</sup> 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.

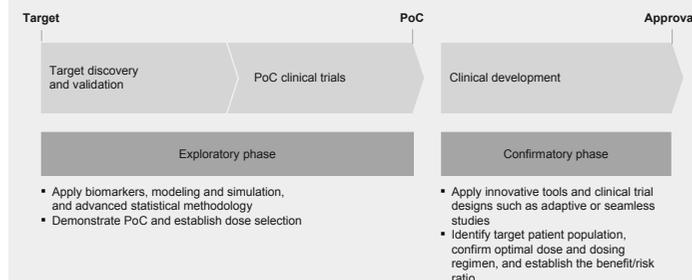
As recognized by the Critical Path Initiative of the US Food and Drug Administration (FDA), novel approaches to clinical trial and program design could have an important role in overcoming these challenges. The traditional approach to drug development separates clinical development into sequential,

*During the exploratory phase of development, this new model uses all available knowledge and tools, including biomarkers, modeling and simulation, and advanced statistical methodology. Trials are designed to determine proof of concept (PoC) and to establish dose selection to a level of rigor 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 goals 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.*

### EXHIBIT 1 A novel model for clinical development



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” (Exhibit 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 are novel tools, including modeling and simulation, Bayesian methodologies, and adaptive designs such as seamless adaptive designs and sample-size reestimation methods (see sidebar 1, “Tools, methods, and designs”). These can ensure the judicious use of limited patient resources, reduce patients’ 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 represent the most favourable benefit/risk ratio.

Here we describe the general issues and methods involved, and show how the tools can be applied in both exploratory and confirmatory 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

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

In the exploratory phase, modeling and simulation can help refine dose selection and study design. Early development studies are conducted with fairly restricted resources (limited duration, sample sizes, and so on), and the use of all available information is thus crucial for effective decision making.<sup>16</sup> 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 prove not to correlate with, or be predictive of, the final outcome. Accordingly, it is important to conduct methodology research in parallel with the development program to establish the correlation between the biomarker and late-stage endpoints or outcomes.

Modeling and simulation approaches can be used to represent dose-response and time-response behavior 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

## 1. Tools, methods, and designs for enhancing development

### Modeling and simulation

*These techniques are a cornerstone of the novel drug-development model. In the exploratory phase, modeling and simulation can help to refine dose selection and study design, and represent dose-response and time-response behavior of safety and efficacy endpoints. In combination with Bayesian methods, these can provide a continuous flow of information across different phases of development. Modeling in early development also enables the use of external information (an important issue in the light of current discussions within the industry about sharing placebo data among 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 footnotes 22 and 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 more effectively to 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, the decision to drop poorly performing trial arms, and the decision to terminate the trial for futility, and make 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

*A seamless design combines, in a single trial, 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 reestimation methods

*These provide the flexibility to increase or decrease the sample size at an interim point in the trial. This is important if there is uncertainty about between-subject variance in the response, or 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.*

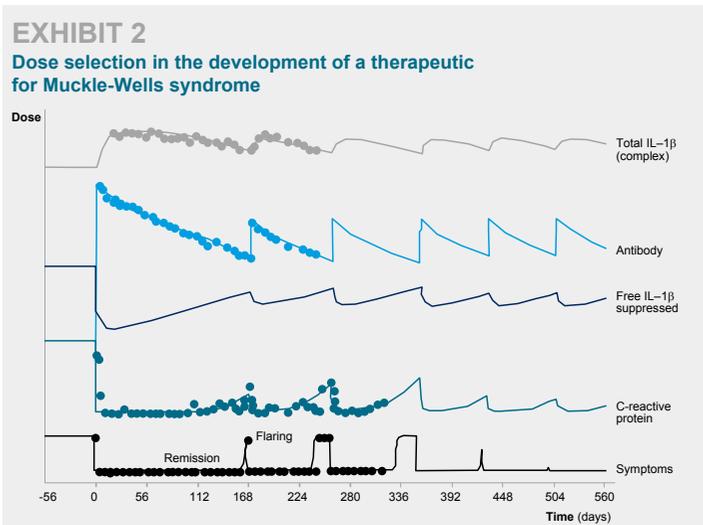
data can be used to construct models and to provide prior information on model parameters. Likewise, the results from a proof-of-concept (PoC) study can be used to form prior distributions for a similar model to be used in a subsequent dose-finding study.<sup>11,12,17,18</sup>

An additional benefit of modeling 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 programs 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 to improve the efficiency of investigation in early development.

#### Modeling and simulation for dose and dose regimen selection.

An important goal of a drug-development program is the selection of a dose and dosing regimen that achieves the target clinical benefit while minimizing undesirable adverse effects. Biological and pharmacological modeling can be useful in this context.<sup>19,20</sup> For example, Novartis has used it 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 (Exhibit 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

*Muckle-Wells syndrome is a rare genetic disorder characterized by fever, urticaria, joint pain, and malaise. A monoclonal antibody against interleukin-1 $\beta$  (IL-1 $\beta$ ), canakinumab, has been developed to treat such an IL-1-dependent inflammatory disease. The antibody is delivered parenterally and binds to free IL-1 $\beta$ , driving it into the inactive complex and leading to remission of symptoms.<sup>21</sup> Total IL-1 $\beta$ , 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-1 $\beta$ , which cannot be measured, must decrease. However, the reduction in free IL-1 $\beta$  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 exhibit, and 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 percent probability that 90 percent of patients will be flare-free for two months).*



and dosing regimen that would achieve the desired response for the majority of patients—in this instance, an 80 percent probability that 90 percent of patients would remain flare-free for two months. The data derived from this modeling exercise allowed for selection of a dosing regimen that was investigated and confirmed in a Phase III trial<sup>21</sup> (clinical data on various dosing intervals provided the raw data for the modeling and simulation exercise that finalized the dose and regimen selection for Phase III). Similarly, modeling 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.

**Bayesian modeling combined with use of external baseline data to improve efficacy and detection of safety signals in early development.** Early development studies for establishing PoC often use small patient cohorts (ten to 20 subjects). These patients are usually observed for a relatively short period (several weeks) to evaluate early efficacy and safety signals, which are frequently measured on a continuous scale. 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 behavior. 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.<sup>22</sup> 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 (MTD) is reached, or unacceptable safety is observed for a given cohort. A new cohort is allowed to start only when acceptable safety signals are verified for all previous doses. At the end of the study, the goal is either to 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 deploying longitudinal modeling approaches to make use of all available information. Furthermore, the utility of PoC studies within drug-development programs can be enhanced by incorporating the information obtained in them directly into later-phase trials.<sup>11,12</sup> Bayesian modeling techniques are particularly useful in implementing these approaches.

#### 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 how to modify aspects of the study mid-trial, in a pre-planned manner, without undermining the validity or integrity of the study.<sup>23-26</sup> Possible adaptations include adjustments to sample size, allocation of treatments,

## 2. Case study: combining proof-of-concept and dose-ranging trials into a single adaptive trial

This example shows how a single adaptive trial can replace two standard trials—PoC and dose-ranging. It also shows 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 Exhibit A. 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 Exhibit B.

When there was a small dose-response effect (scenarios two and three), the power of the adaptive design was about double that of the standard design. When the effect size was modest (scenarios four and five), the power was increased to almost 100 percent. When effect sizes were large (scenarios six and seven), the power was almost 100 percent for both adaptive and standard designs.

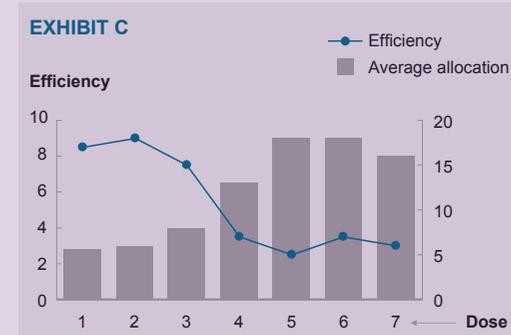
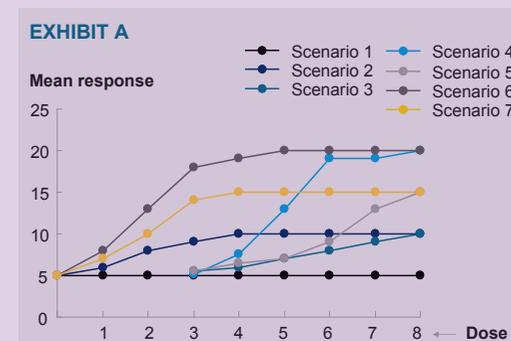
For the same total sample size, the adaptive combined PoC-dose-finding trial is more efficient than the two standard trials in estimating the response at every dose (Exhibit C). The continuous curve shows the efficiency of the adaptive design relative to the standard dose-ranging design for scenario seven. 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 four, five, and six), the adaptive design is three times more efficient. In Exhibit 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.

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 programs 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 (see sidebar 2, “Case study”). This has the advantage of reducing start-up costs and the time between trials, while potentially increasing statistical power and improving estimates of dose response. Adaptive designs can also enable trialists to work with more candidate doses without increasing the sample size. This is important to reduce the risk of failure in confirmatory trials, in which, it is estimated, 45 percent of Phase III programs industry-wide do not have the optimum dose.<sup>3</sup> Adaptive dose-ranging studies are discussed further in sidebar 3, “Adaptive dose finding.”<sup>27, 28</sup>

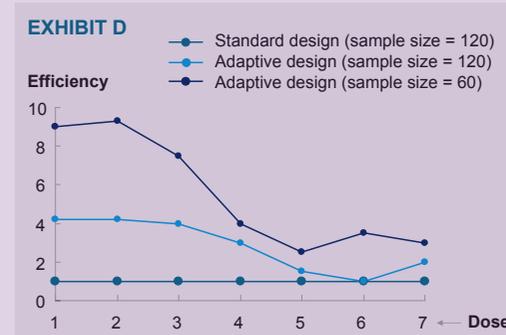


The above results are true irrespective of which of the seven scenarios reflects the true dose-response curve. For all seven scenarios for the same sample size, the efficiency of the adaptive design was about five times that of the standard design over all doses. It was three times that of the standard 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.

Successful implementation of adaptive trial designs requires a number of things. 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 upfront 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

Scenario	Power of standard PoC trial <sup>1</sup>	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

<sup>1</sup> Sample size = 30



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 find efficiently and confidently the actual effect size of 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 programs, and lower attrition rates. A number of novel approaches to confirmatory development that can contribute to fulfilling this promise are highlighted below.

**Seamless adaptive designs.** Efficiency can be increased through the use of seamless adaptive designs, which aim to combine objectives traditionally addressed in separate trials in a single trial.<sup>25,29</sup> An 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 main potential advantages of seamless adaptive designs: they shorten the clinical-development program by eliminating the time lag between Phase II and III trials; they lead to greater efficiency in the use of data from both stages, which might mean that fewer patients are required to obtain the same quality of information; and they enable the earlier acquisition of long-term safety data, gathered through continued follow-up of patients from the first stage.<sup>25,29</sup>

The designs are not suitable for all drug-development programs. Feasibility considerations include the length of follow-

up time for the endpoint used for selection compared with duration of enrollment. Shorter follow-up will be more conducive to their use, whereas a relatively long endpoint follow-up period will tend to militate against. Development programs that do not involve complex treatment regimens might thus lend themselves better. Drug supply and drug packaging will be expected to be more challenging in this setting.

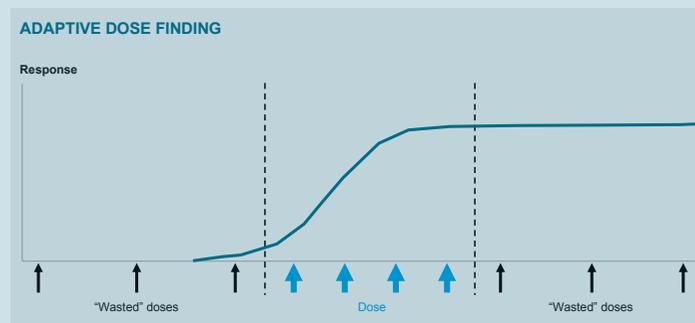
A number of logistical and regulatory actions must be taken to avoid compromising an adaptive trial. First, the actual algorithm for deciding the adaptation to implement must be specified in advance. This is usually accomplished by creating a charter for the independent data-monitoring committee charged with 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 a trial's integrity, the processes by which interim data are examined and selection decisions are made and implemented must be considered 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 a sponsor's perspective might be relevant and for which the sponsor traditionally has been responsible, raising the question of its involvement in the monitoring process. A distinction is sometimes made between seamless adaptive designs that are inferentially seamless and those that are operationally seamless. In the former, which we describe here, the main analysis uses data from both stages of the trial. In the latter, the

### 3. Adaptive dose finding

*In an adaptive dose-finding study, the dose assignment to the next subject, or next cohort of patients, is based on the responses of previous subjects and 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 might not adequately represent the dose-response relationship and many patients will be allocated to "non-informative" doses (wasted doses), as shown in the exhibit. In adaptive dose-finding, the strategy is initially to 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.<sup>27,28</sup> 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 enables the rapid communication of responses from trial sites to a central, unblinded analysis center 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.*

final analysis uses only data from patients enrolled after the selection decision. This might allow a broader investigation of the first-stage data involving the sponsor's personnel, while alleviating concerns about the trial's 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 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, modeling and simulation will probably have an 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: "naïve" comparisons of control versus the selected treatment that do not account for the

#### 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  $\delta$  (1,000 subjects). Now, suppose that the true underlying value of  $\delta$  is 0.15. In that case, a sample size of 2,000 subjects would be required to power the trial adequately to detect this difference. The difficulty is, of course, that the true underlying value of  $\delta$  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 powered only 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 were repeated with a larger sample size, a potentially efficacious treatment would be unnecessarily and unfortunately discarded. If the trial were to be repeated with the reestimation 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 three years) would have been spent without much benefit other than to gain 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.

design will not be appropriate. Finally, the aptness 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 reestimation within a confirmatory trial (Phase III).** Sample size reestimation (SSR) provides a mechanism for the appropriate use of the information obtained during a confirmatory study to inform and adjust the necessary sample size going forward.<sup>30,31</sup> This process increases the 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 first to estimate the underlying treatment effect of the primary endpoint based on available prior information. The parameter  $\delta$  denotes the true underlying difference between the

treatment and control arms with respect to the primary endpoint. Even though the true value of  $\delta$  is unknown, the trial investigators will usually have in mind a specific value,  $\delta_{\min}$ , which represents the smallest clinically important delta (SCID) for this trial. Next, the trial designers will decide the sample size that can detect values of  $\delta$ , based on prior information, that exceed the SCID with good power. The standard deviation  $\sigma$  (between subject variability) is a “nuisance parameter” the true value of which must be estimated in order to proceed with the sample size calculation.

The SCID can be often pre-specified from purely clinical arguments, whereas the actual effect size is unknown. Thus 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

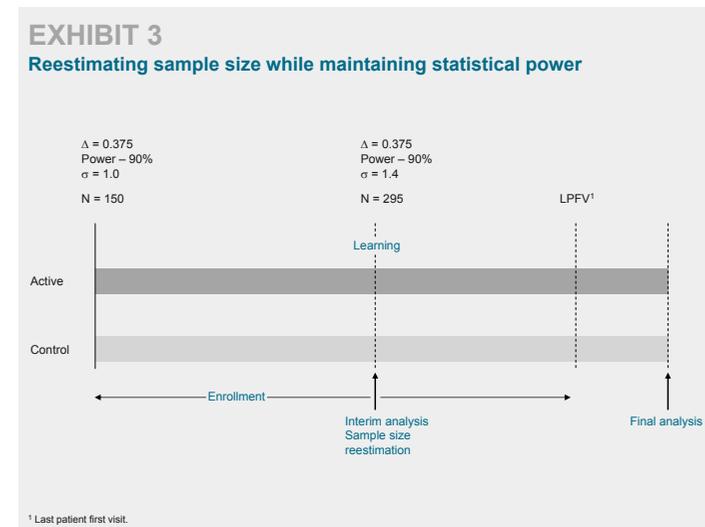
methodology has several drawbacks. If the actual effect is substantially larger than the SCID, a smaller sample size would have sufficed to attain adequate power.<sup>32</sup> Sponsors will not often risk significant resources on trial sizes based on SCID assumptions that would lead to larger trials than the current “best guess” about the actual effect size (see sidebar 4, “Issues with the standard approach”). 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 thus have a high chance of failure.

One approach to solving the problem of uncertainty about  $\delta$  is to design and execute further 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  $\delta$  and  $\sigma$  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 deciding true program efficiency and development timelines. This might therefore not be the most efficient way to proceed from the viewpoint of the entire clinical-trial program.

**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: where there is significant uncertainty about  $\sigma$ ; or where 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 its best guess of the effect size.<sup>29</sup>

This is a hypothetical example of a study in which sample size reestimation owing to uncertainty about  $\sigma$  led to



an increase in sample size to ensure 90 percent 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 was as originally predicted, an increase in sample size to 295 patients would be required to maintain 90 percent power. Without the sample size reestimation, the power at the final analysis would be only 64 percent and there would be greater risk of a failed trial.

## 5. Group-sequential and adaptive designs for sample size reestimation

### Group-sequential design

Suppose that the sponsor is unsure of the true value of  $\delta$ , 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  $\delta$  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 upfront commitment is made to enrol patients up to this maximum sample size. However, if the true  $\delta$  exceeds the SCID, the trial might terminate earlier with high probability by crossing an early stopping boundary at an interim analysis.

Returning to the example discussed in Sidebar 4, suppose that the sponsor decides to make an upfront 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 percent power to detect a difference as small as  $\delta = 0.1$  (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  $\delta$  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 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 upfront commitment, particularly when the only currently available data on  $\delta$  come 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 itself. This is the motivation for the adaptive design considered next.

### The adaptive design

The group-sequential design described above is characterized by pre-specifying a maximum sample size upfront and terminating earlier if the true  $\delta$  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 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

SSR usually involves the choice of a suitable initial sample size, including one or more interim analyses at which the sample size will be reassessed.<sup>30</sup> 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 commitment to a large upfront sample size and cuts back

if the accruing data suggest that the large sample is not needed. The adaptive SSR strategy proceeds in the opposite direction, starting out with commitment to a smaller initial sample size but with the option to increase it should the accruing data suggest that an increase is warranted<sup>30-33</sup> (see sidebar 5, “Group-sequential and adaptive designs”).

### Extending the methodology to unknown $\sigma$ .

Although the group sequential and adaptive SSR methods were presented under the

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 and in shortening the drug-development time.

In our example, we stated that the SCID was 0.1. Supposing that the sponsor believes that the true  $\delta = 0.2$ —that is, twice the size of the SCID. If this is indeed the case, then a total sample size of 1,000 patients will have 89 percent power at a one-sided 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

Interim estimate, $\delta$	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 (reduction)
0.15	72	1166 (increase)
0.125	51	1757 (increase)
0.1	30	2990 (increase)

data at the mid-point of the trial, after 500 patients are evaluable for response. If the estimate of  $\delta$  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  $\delta$ . Either the value specified at the initial design stage or the value estimated from the interim data can be chosen. In this example, we use the interim estimated value of  $\delta$  for evaluating conditional power. The table below displays conditional power for various estimated values of  $\delta$  at the interim time point, along with the total sample size needed to achieve 80 percent conditional power at the final analysis. The entries in the table assume that  $\sigma = 1$ . Note that the final sample size required to achieve 80 percent conditional power could increase or decrease from the initially planned sample size of 1,000.

assumption that the standard deviation  $\sigma$  is known, they apply equally for the case of unknown  $\sigma$ .<sup>30-32</sup> One can start out with an initial estimate of  $\sigma$  and a corresponding sample-size estimate. Then, following an interim analysis, one can reestimate this nuisance parameter, enter the updated estimate into the equation, and recompute the sample size. An example is given in Exhibit 3.

There are two ways to obtain the new

sample size in the situation of unknown  $\sigma$ : blinded and unblinded. In the instance of blinded sample size reestimation, the sponsor uses pooled data to estimate  $\sigma$ . 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 reestimation, 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 might 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 the light of efficacy data, do not need to have defined criteria. The trial then proceeds either until it is terminated early for efficacy on the basis of the pre-defined criteria having been reached, or until the planned maximum information (sample size or number of events) is reached.

### Tackling the challenges of adaptive 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 take time too. 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, both across trial sites to the central unblinded analysis centre, and of dose assignments to trial sites; and a flexible drug-supply process. Appropriate models, which reliably characterize the longitudinal behavior 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 would be to use “adaptive modeling” approaches. This is a topic for further research.

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 data sets can be difficult because of the 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 of modeling and simulation, into clinical-development programs.

A key barrier to the implementation of tools and techniques which advance the quality, timeliness, and efficiency of drug development is the ability to work across disciplines and among 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 must be defined and disseminated. Second, guidelines and case studies for assessing situations in which tools should—or should not—be applied must be developed and disseminated. Third, there is a need for a methodology for dialogue with regulatory authorities to facilitate discussion of clinical strategies which use these tools and address potential constraints. Finally, it will be crucial to identify specific solutions to address obstacles and objections that inhibit the adoption of modern tools and adaptive study designs.

This article, with supplementary information, was originally published by *Nature Reviews Drug Discovery*, and can be found for citation purposes online at [www.nature.com/reviews/drugdisc](http://www.nature.com/reviews/drugdisc) (doi:10.1038/nrd3025). Please direct scientific correspondence to John Orloff ([john.orloff@novartis.com](mailto:john.orloff@novartis.com)). Requests of a general nature can also be made to the McKinsey Trial Design Team ([matthias\\_evers@mckinsey.com](mailto:matthias_evers@mckinsey.com)).

John Orloff, Jose Pinheiro, Michael Branson, Paul Gallo, and Donald Stanski are at Novartis Pharmaceuticals Corporation; Frank Douglas is at the Kauffman Foundation; Susan Levinson is at The Strategic Choice; Pravin Chaturvedi is at Napo Pharmaceuticals; Ene Ette is at Anoxis Corporation; Gigi Hirsch, Nitin Patel, Sameer Sabir, Stacy Springs, and Howard Golub are at the Center for Biomedical Innovation, Massachusetts Institute of Technology; Cyrus Mehta is at Cytel Statistical Software and Services; and Matthias Evers ([matthias\\_evers@mckinsey.com](mailto:matthias_evers@mckinsey.com)), Edd Fleming ([edd\\_fleming@mckinsey.com](mailto:edd_fleming@mckinsey.com)), Navjot Singh ([navjot\\_singh@mckinsey.com](mailto:navjot_singh@mckinsey.com)), and Tony Tramontin ([tony\\_tramontin@mckinsey.com](mailto:tony_tramontin@mckinsey.com)) are at McKinsey & Company.

We would like to acknowledge W. Dere (Amgen), S. Cummings (UCSF), A. Lee (Pfizer), and E. Berndt (MIT-CBI) for significant contributions to discussions leading to this manuscript.

- <sup>1</sup> Booth, B. and Zimmel, R., "Prospects for productivity," *Nature Reviews Drug Discovery*, 3, 451–456 (2004).
- <sup>2</sup> Gilbert, J., Henske, P., and Singh, A., "Rebuilding big pharma's business model," *In Vivo* 21, 1–4 (2003).
- <sup>3</sup> Kola, I. and Landis, J., "Can the pharmaceutical industry reduce attrition rates?" *Nature Reviews Drug Discovery*, 3, 711–716 (2004).
- <sup>4</sup> Budget US Government, App., FY 1993–2003.
- <sup>5</sup> *Parexel's Pharmaceutical R&D Statistical Sourcebook 2002/2003* (Parexel International, Waltham, US, 2003).
- <sup>6</sup> Adams, C. P. and Brantner, V. V., "Spending on new drug development," *Health Economics*, February 26, 2009 (doi: 10.1002/hec.1454).
- <sup>7</sup> DiMasi, J. A., Hansen, R. W., and Grabowski, H. G., "The price of innovation: new estimates of drug development costs," *J. Health Econ.* 22, 151–185 (2003).
- <sup>8</sup> Adams, C. and Brantner, V. V., "Estimating the cost of new drug development: is it really \$802 million?" *Health Aff.*, 2, 420–428 (2006).
- <sup>9</sup> *Pharmaceutical R&D Factbook 2007* (CMR International, London, UK, 2007).
- <sup>10</sup> KMR General Metrics Study (KMR Group, Chicago, US, 2007).
- <sup>11</sup> Sheiner, L. B. and Steimer, J. L., "Pharmacokinetic/pharmacodynamic modeling in drug development," *Ann. Rev. Pharmacol. Toxicol.* 40, 67–95 (2000).
- <sup>12</sup> Lalonde, R. L. et al., "Model-based drug development," *Clin. Pharmacol. Ther.* 82, 21–32 (2007).
- <sup>13</sup> Breimer, D. D. and Danhof, M., "Relevance of the application of pharmacokinetic-pharmacodynamic modeling concepts in drug development. The 'wooden shoe' paradigm," *Clin. Pharmacokinet.* 32, 259–267 (1997).
- <sup>14</sup> Danhof, M., Alvan, G., Dahl, S. G., Kuhlmann, J., and Paintaud, G., "Mechanism-based pharmacokinetic/pharmacodynamic modeling—a new classification of biomarkers," *Pharm. Res.* 22, 1432–1437 (2005).
- <sup>15</sup> Holford, N. H. G., Kimko, H. C., Monteleone, J. P. R., & Peck, C. C., "Simulation of clinical trials," *Ann. Rev. Pharmacol. Toxicol.* 40, 209–234 (2000).
- <sup>16</sup> Miller, R. et al., "How modeling and simulation have enhanced decision making in new drug development," *J. Pharmacokinet. Pharmacodyn.* 32, 185–197 (2005).
- <sup>17</sup> Berry, D. A., "Bayesian statistics," *Med. Decis. Making* 26, 429–430 (2006).
- <sup>18</sup> Müller, P., Berry, D. A., Grieve, A. P., Smith, M. and Krams, M., "Simulation-based sequential Bayesian design," *J. Stat. Plan. Inference* 137, 3140–3150 (2007).
- <sup>19</sup> Goggin, T. et al., "Modeling and simulation of clinical trials: an industrial perspective," *Simulation for Designing of Clinical Trials* (eds. Kimko, H. C. and Duffull, S. B.), 227–224 (Marcel Dekker, New York, US, 2002).
- <sup>20</sup> 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).
- <sup>21</sup> Lachmann, H. J. et al., "Use of canakinumab in the cryopyrin-associated periodic syndrome," *N. Engl. J. Med.* 360, 2416–2425 (2009).
- <sup>22</sup> Berry, D. A., "Bayesian clinical trials," *Nature Reviews Drug Discovery* 5, 27–36 (2006).
- <sup>23</sup> Berry, D. A., "Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis," *Clin. Trials* 2, 295–300 (2005).
- <sup>24</sup> Berry, D. A., "Adaptive trial design," *Clin. Adv. in Hematol. Oncol.* 5, 522–524 (2007).

- <sup>25</sup> 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).
- <sup>26</sup> 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).
- <sup>27</sup> Bornkamp, B. et al., "Innovative approaches for designing and analyzing adaptive dose-ranging trials," *J. Biopharm. Stat.* 17, 965–995 (2007).
- <sup>28</sup> Weir, C. J., Spiegelhalter, D. J., and Grieve, A. P., "Flexible design and efficient implementation of adaptive dose-finding studies," *J. Biopharm. Stat.* 17, 1033–1050 (2007).
- <sup>29</sup> Bretz, F., Schmidli, H., König, F., Racine, A., and Maurer, W., "Confirmatory seamless Phase II/III clinical trials with hypotheses selection at interim: general concepts," *Biom. J.* 48, 623–634 (2006).
- <sup>30</sup> Mehta, C.R. and Patel, N.R., "Adaptive, group sequential and decision theoretic approaches to sample size determination." *Stat. Med.* 25, 3250–3269 (2006).
- <sup>31</sup> Chuang-Stein, C., Anderson, K., Gallo, P., and Collins, S., "Sample size reestimation: a review and recommendations," *Drug Inf. J.* 40, 475–484 (2006).
- <sup>32</sup> Golub, H. L., "The need for more efficient clinical trials," *Stat. Med.* 25, 3231–3235 (2006).
- <sup>33</sup> Tsiatis, A. A. and Mehta, C., "On the inefficiency of the adaptive design for monitoring clinical trials," *Biometrika* 90, 367–378 (2003).



## Using IT to speed up clinical trials

The first wave of IT improvements solved some problems, but others remain. Now it is time for a more comprehensive approach.

Sam Marwaha, Samir Patil, and Navjot Singh

Clinical trials are costly and complex undertakings for pharmaceutical companies. A trial can cost hundreds of millions of dollars and require the coordination of many patients, physicians, and regulatory agencies for up to seven years. The stakes are high, because companies spend about 10 percent of their operating budgets on these trials and, more importantly, can derive huge benefits from being first to market with a new type of drug. But they face intense competition from similar compounds and from generic drugs once their patents run out. As patents are issued before a drug goes into clinical trials, the faster a trial goes, the longer the manufacturer enjoys a monopoly until generic versions can be sold. Thus, a streamlined and speedy trials process can have a significant impact on the economics both of new drugs and of existing drugs that may be approved for additional uses after further clinical studies.

Over the past decade, pharmaceutical companies have introduced a number of initiatives to boost the productivity of trials. They have rolled out electronic data capture (EDC) systems, allowing patients and researchers to enter trial information directly in electronic diaries or online systems. They have adopted new scientific approaches, such as Bayesian techniques<sup>1</sup> that enable them to refine

their designs for trials from one stage to the next. They have improved their technological capabilities—especially connectivity—at clinical locations, enabling trial managers to keep tabs on the retention of patients and the progress of trials. And they have enhanced their ability to conduct trials worldwide by vastly increasing the pool of researchers and patients. They have also adopted more disciplined procedures for managing trials, in some cases borrowing techniques (including “stage gates,” which set firm deadlines for gathering data and refining the goals of subsequent stages) from product design.

Overall improvements in the performance of trials remain elusive, however, especially in the early stages. There are several reasons. First, many companies fail to coordinate trials across their organizations as well as they could, and the lack of cross-trial transparency can create delays when different trials compete for scarce resources. Second, many organizations have yet to embrace reusability by streamlining their approach to the design of trials; for example, certain components of the forms that guide researchers in trials could be shared and reused in other trials. Third, although EDC systems have substantially reduced the time required to gather data—to two weeks, in some cases, from 20—the first

generation of systems was not always flexible or reliable enough to suit investigators' needs. Fourth, verifying electronic data to ensure that they correlate with physical data (such as laboratory test reports) remains time consuming. Finally, the productivity measures some companies have implemented apply only to parts of the organization and thus fail to capture all of the potential benefits.

In recent years, some leading companies have begun to ease the bottlenecks and raise productivity by revisiting IT systems to streamline the process further.

Such efforts are most successful when companies take a comprehensive, "clean-sheet" approach, integrating their systems in a redesign of the whole trials program (Exhibit 1). The application of these principles has speeded up some trials by 10 percent or more. In addition, a number of companies have applied principles from lean manufacturing to these information flows, improving their throughput and reducing waste by redesigning their clinical trials processes, staff responsibilities, and technology. Certain companies have enhanced their speed, quality, and costs by focusing on getting the right data the first time, managing work flows to reduce bottlenecks, and making data more transparent across the entire clinical process.

Through our work with these companies, we have identified four areas in

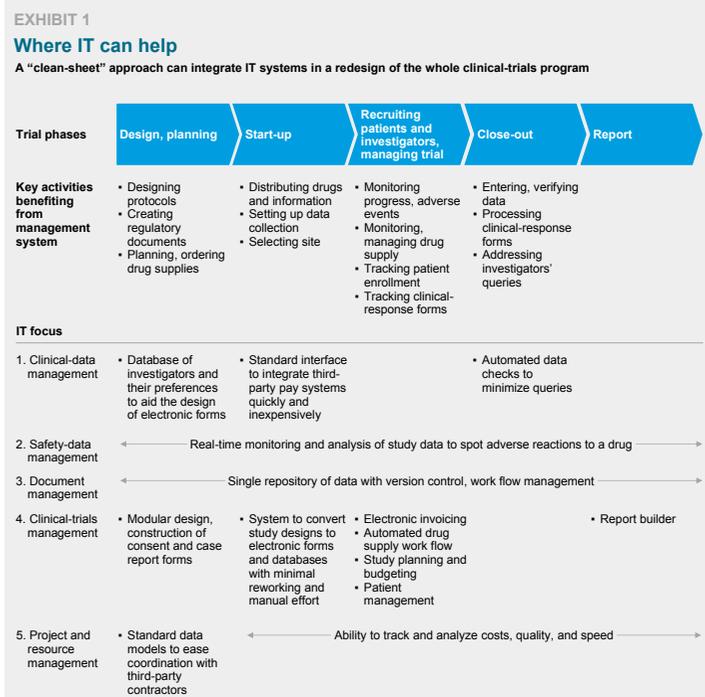
information and process design that are ripe for improvements to make clinical trials faster and cheaper.

**Planning.** Integrated planning of clinical trials improves the way resources (for example, clinicians and biostatisticians) are allocated to clinical teams.

**Tools.** Modular designs and reusable components enable the rapid creation of electronic case-report forms.

**Use of tools.** Physicians can be helped to use EDC tools more effectively through the provision of training and customer support, and the introduction of a standard interface.

**Integrated architecture for information systems.** This will make trials visible to management from start to finish, and thus more efficient, by eliminating bottlenecks.

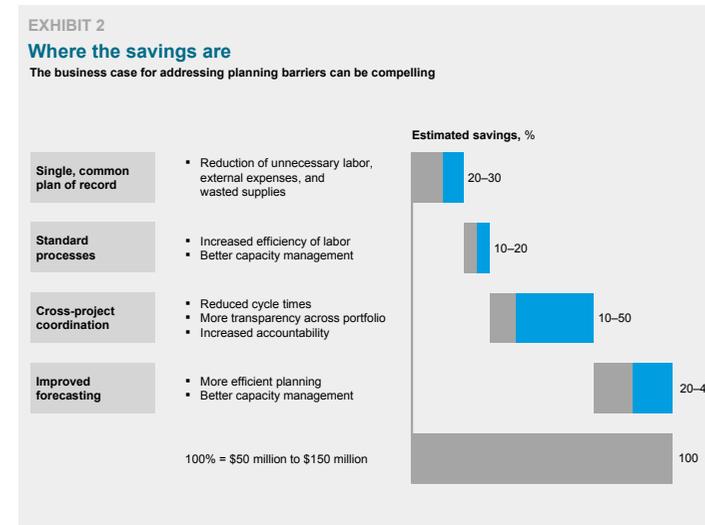


## Developing an integrated plan

Pharmaceutical companies design trials within individual therapeutic areas, but trials in all of them draw from a pool of shared functions. Coordinating these resources is not easy given the dynamic nature of the trials portfolio, the challenge of creating the right cross-functional teams from the shared pool, and the difficulties of recruiting investigators and patients. To complicate matters, program managers for trials often try to reserve more resources than they need in order to ensure that the actual requirements will be available. What is more, when trials are canceled, newly freed-up resources often are not redistributed efficiently to ongoing efforts. These factors can make the use of a company's resources less than optimal, limiting its ability to initiate new studies or to complete ongoing ones quickly. Developing an integrated plan across therapeutic areas to address these issues can cut costs and speed up the completion of trials: one large company identified potential savings of \$50 million to \$150 million from improving the way it allocates resources across projects (Exhibit 2).

The transformation from *ad hoc* planning to an integrated approach requires a comprehensive program that addresses the cultural, process, and systems barriers to change. Five elements are vital.

1. Support from senior managers, who must insist, very visibly, on the use of standardized processes and systems.
2. A common process, that all project managers follow, for formalizing the way the company creates, approves, controls, executes, and ends trials.
3. A flexible approach to planning that enables project teams to plan at whatever level is clear to them—an entire program, a project within a larger program, or a bundle of tasks within a project.
4. Forecasts of resource use based directly on project plans.
5. Common standards for detailed descriptions of work activities that will help resources to be allocated more efficiently and specifically.



Each of these elements has implications for information technology. IT platforms for enterprise project management<sup>2</sup> help a company to develop and manage a more integrated approach. In addition to linking available resources and trial schedules automatically, the platform serves as a repository of design templates for trials, enforces standards, and helps to improve work flows.

## Designing trials rapidly through modularization

One result of integrated planning across therapeutic areas is that companies are beginning to standardize their processes and tools for designing and recording a trial, and for guiding the investigators who conduct it. Designing a trial begins with setting its aims (that is, what the researchers hope to prove about a new drug) and describing the type of patients and physicians it will require. The designers then work with colleagues—within the company and in partner organizations such as contract research bodies that help coordinate trials—to convert these aims into a series of documents, including the patient consent form, the list of laboratory tests to be conducted, and the case report form that physicians fill out during the trial.

Although there is a robust culture of independence among designers of trials, several factors are encouraging standardization and reuse. The first is the recognition that integrated planning helps make better use of the organization's resources. Creating reusable modules that can be applied in a number of studies (for example, a standardized approach to designing and collecting data about a drug's safety and any adverse effects) saves time in subsequent designs. Senior managers, recognizing this opportunity, are often strong advocates of the shift to modular design.

Second, the advent of EDC creates an incentive for designers of trials to modularize the process in order to minimize the cost of developing and integrating electronic case reports for each one. First-generation EDC systems created some unforeseen fixed costs. For example, before they were introduced,

information was collected manually, so it was easy for the person recording it to spot and correct obvious errors early in the process. Programming EDC systems to identify these errors, through a series of checks, is time consuming and generates additional start-up costs for trials. Standardized error-checking modules, by contrast, allow such checks to be built once and reused many times.

Third, collecting data in a standard form makes it easier to perform analyses early in a trial. This in turn enables project managers to see initial findings more quickly and to make interim decisions about the rest of the trial.

## Tailoring electronic data collection to investigators' needs

Once the design is completed, the manager (often a contractor) must recruit and train investigators to undertake the trials. In most cases, the physicians who act as investigators have to balance competing demands: the extra effort required to carry out cutting-edge research with the need to devote sufficient time to manage a practice and care for patients. Conducting efficient clinical research trials is therefore a primary concern for investigators. Any improvement in efficiency will help pharmaceutical companies to sign up leading practitioners.

In recent years, investigators have become increasingly frustrated by EDC systems, which, they say, create more work than the old-fashioned pen-and-paper method did; sometimes, for the sake of reliability they have had to keep data both ways. Doctors and nurses also complain that they lack adequate training in the systems—a problem complicated by the fact that every company uses its own proprietary systems.

The pharmaceutical industry could improve its ability to sign up physicians by making data-collection systems more friendly to them in several ways. First, companies should gather and track physicians' preferences in order to understand how they want to use electronic forms. They should then ensure that the systems are flexible enough to accommodate these preferences. If some leading researchers insist on staying with pen and paper, designers of trials must make sure that there is a solid process to convert the data to standard electronic forms (perhaps in a low-cost remote location).

Likewise, companies should standardize the way they collect data not only within their organizations but also across the industry (see sidebar, "A Sabre for pharmaceuticals?"). This move will shorten the learning curve for physicians as they will have to master the electronic-entry system once only.

Finally, companies should redefine an important factor in trials: the role of clinical-research associates. Currently, associates mainly monitor programs. Instead, they

should become more active problem solvers who manage the relationship between the organization conducting the trial and the investigators. As part of this shift, their incentives should reflect not just the number of visits they make to sites but also their speed in recruiting patients and the quality of the data gathered.

## Streamlined work flow and maximum trial visibility

The fragmented IT systems and data architecture<sup>3</sup> of most pharmaceutical companies make it extremely difficult for managers to track productivity throughout a trial. Managers might be able to reduce bottlenecks in the availability of patients, for example, if they could see the patient-recruitment "pipeline." But although they have access to the clinical-trials management system, information on patient recruitment is usually kept in a separate EDC system—or even on paper.

A single, real-time view of all the relevant data on trials could help

### A Sabre for pharmaceuticals?

*Just as American Airlines' Sabre system accelerated the adoption of electronic ticketing in the airline industry by offering a standard booking interface across many air carriers, so too could a standard system speed up the adoption of electronic data capture in pharmaceutical trials. Although investigators and other researchers currently have to learn a new system for every pharmaceutical company or clinical research organization with which they work, change is on the way.*

*One initiative, sponsored by the US National Institutes of Health, is called the Clinical Research Information Exchange. It aims to bring commercial and academic researchers together in a system that automates and standardizes registration forms for trials. However, because industry-wide efforts often move slowly, other organizational forms may emerge to drive the next level of standardization. An open-source foundation could develop a core set of standard tools coupled with a for-profit ecosystem of service providers that would supply, maintain, and configure such systems as part of their work with pharmaceutical companies. A small consortium of pharmaceutical companies, working in the same therapeutic areas, might coordinate their efforts in electronic data capture. Finally, a commercial organization, financed by venture capital or private equity, could sign up large pharmaceutical companies as its lead customers to drive the standardization of systems across the industry.*

companies to overcome these barriers. IT suppliers hope to provide more integrated systems, but the task may take a while. In the meantime, therefore, companies should continue to take a best-of-breed approach and invest in middleware tools that help to exchange data among systems. A “dashboard” that displays key metrics in a spreadsheet or some other easy-to-view form can also help managers to keep tabs on the bigger picture.

Even across a number of systems, companies must work to streamline and unify the way they manage work flows. Automated notification—such

as alerts to warn when data entry has been delayed—is one example of how IT can help managers to coordinate projects to ensure that bottlenecks do not occur. Alerts could also help to trigger coordination across functions to keep processes moving without delays.

Finally, a standard interface can help companies to analyze and compile data received from a number of different partners that manage studies within a single program. Currently, each third party might use its own data standard. Standardizing within a company (and across the industry) could make it easier to exchange data and speed up trials.

**Sam Marwaha** (*sam\_marwaha@mckinsey.com*) is a director in McKinsey’s New York office, where **Navjot Singh** (*navjot\_singh@mckinsey.com*) is a principal. **Samir Patil** is a McKinsey alumnus.

<sup>1</sup> Bayes’ theorem explains how to update or revise beliefs in the light of new evidence.

<sup>2</sup> Enterprise project-management platforms are large project-management systems that enable companies to manage portfolios of projects more efficiently across a whole organization. Depending on a company’s needs and capabilities, such a system might be either a custom package developed by a single vendor or a combination of “best-of-breed” applications and tools.

<sup>3</sup> At many pharmaceutical companies, systems for developing trials are fragmented into five clusters: systems to manage trials, to capture data, to manage data safely, to manage documents, and to manage projects and resources.

## Quality gates: raising the standard of pharmaceutical technical development

Good technical development can improve quality, lower costs, and get products to market more quickly, yet few pharmaceutical companies manage it well. Quality gates can transform the process.

Michele Holcomb, Martin Møller, and Paul Rutten

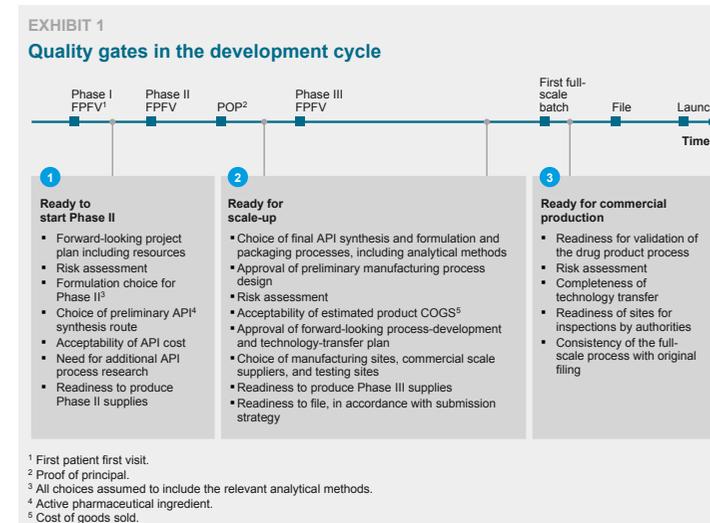
New drugs have to be clinically effective. But to be commercially successful, they also have to be mass produced safely and reliably in a form that ensures a consistent therapeutic effect. This technical-development process<sup>1</sup> absorbs between 15 and 30 percent of the development costs of a typical small-molecule drug and even more in biologicals. Indirectly, it affects 50 percent of those costs if one includes, say, delivery of the supplies needed to run clinical and non-clinical trials.<sup>2</sup>

Good technical development can improve the quality of products and manufacturing processes,<sup>3</sup> lower costs, and get products to market more quickly.

Innovations in technical development can also open up new therapeutic uses, extend the life of a patent, and help products win a distinctive market position. Yet many companies still struggle to manage their technical-development processes well. Indeed, we believe that the pharmaceutical industry could raise its profits by up to 20 percent if companies systematically pursued opportunities to reduce the costs and increase the revenue of technical development. A management tool known as the quality gate is key to reaping such rewards.

### Identifying risk early

Technical development is a complex and cross-functional activity that must overcome manufacturing challenges, take account of quality assurance, intellectual property protection, and regulatory compliance, and address scientific and medical issues. The traditional management approach is reactive: when problems occur, cross-functional groups come together to find solutions. But the end result is often higher costs or launch delays.



Part of the problem has been that R&D managers give scant attention to technical development, often regarding it as an enabler rather than a driver of value creation and focusing solely on development speed. Even when managers do try to assess technical development, the highly specialized and science-based nature of the work makes evaluation difficult. The first indication of trouble often comes only when the technical-development team misses a critical project milestone.

Some companies are starting to do things differently. Quality by Design is a proactive, risk-based approach to technical development that helps identify potential problems before they occur, and puts mitigation strategies in place to ensure that the interests of all relevant functions are served quickly, smoothly, and cost effectively. At the heart of this new approach are quality gates, used extensively in industries such as the automotive sector to manage the complex, costly, and time-consuming development process, and applied with great success in clinical development in the pharmaceutical industry. Many pharmaceutical companies claim to be using quality gates to manage technical development too, but few do so rigorously enough to reap the potential rewards. Used well, quality gates can transform the technical-development process.

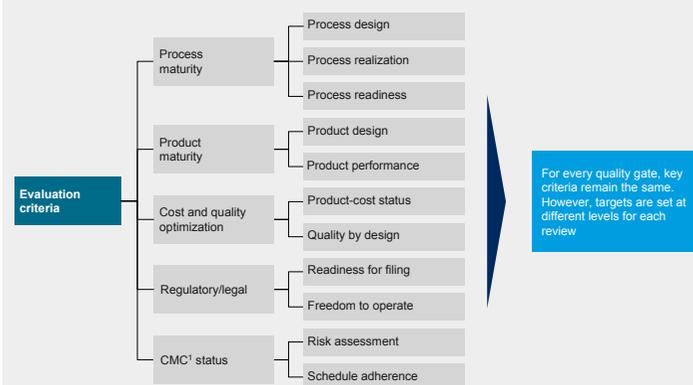
Quality gates provide a mechanism to enforce regular, cross-functional evaluation of development risks. They measure the progress of the technical-

development project against a clear target state with attention not only to development speed, but also to future cost (or process robustness) and potential quality or regulatory issues. Any deviations are identified quickly and appropriate remedial action is agreed. Importantly for pressed managers, assessing a project at a quality gate is not a lengthy process. A typical quality gate review takes less than half a day, and some companies use as few as four gates during the clinical-development cycle of a new drug to assess whether it is ready for clinical trials, ready to start Phase II, ready for scale-up, and ready for commercial production (Exhibit 1).

### How quality gates work

Quality-gate meetings are attended by key members of the project team and a management committee, often consisting of functional heads of chemistry, manufacturing, and controls (CMC), production, quality assurance, and regulatory affairs, and preferably also a senior business leader such as head of R&D. All functions need to be represented to ensure the meeting focuses on time,

EXHIBIT 2  
Evaluating progress



¹ Chemistry, manufacturing, and controls.

EXHIBIT 3  
Traffic light coding

Evaluation	Definition	Consequences
	<ul style="list-style-type: none"> <li>There are significant gaps in the work OR</li> <li>Significant deviations from targets are expected OR</li> <li>Major risks exist which cannot be mitigated</li> </ul>	<ul style="list-style-type: none"> <li>Corrective action defined</li> <li>Necessary escalation taken in a timely manner</li> <li>Strict follow-up defined</li> </ul>
	<ul style="list-style-type: none"> <li>Current status deviates significantly from targets, but actions are in place to close the gap OR</li> <li>There are significant risks, but actions are in place that mitigate the risks</li> </ul>	<ul style="list-style-type: none"> <li>Technical development and clinical project leadership acknowledge deviations, risks, and action plan</li> <li>Arms-length monitoring of progress defined</li> </ul>
	<ul style="list-style-type: none"> <li>No/little deviation from target status at current time</li> <li>Limited risks</li> </ul>	<ul style="list-style-type: none"> <li>No corrective action needed</li> </ul>

quality, and costs. Meetings should be designed using three basic principles.

1. They should be **purposeful and proactive**. This means they should add value for all participants: they are not a reporting exercise. The objective is to smooth progress towards project goals, both in the near and longer term.
2. They should be **simple**. Each gate should require the minimum necessary preparation.

The process should be easy to follow and to communicate. Evaluation and target criteria should be clear and readily understood.

3. They should be **consistent and predictable**. Gates should use the same evaluation criteria for all projects and progress should be tracked throughout the project using consistent metrics. In this way, information is easily accessible, comparisons can be made between projects,

and those who perform the quality-gate reviews gradually develop a valuable sense of pattern recognition.

Quality gates call for a collaborative working style, and participants must commit themselves to raising any problem as soon as possible. Discussions should be open and constructive, concentrate on finding solutions quickly, and culminate in defining next steps, individual responsibilities, and deadlines.

EXHIBIT 4  
Warning signs

Evaluation criteria			Quality gate 2 – ready for scale-up			
			API	Formulation	Packaging	
1. Process properties						
1.1	Process design	1.1.1	Process options	Final API synthesis route has been defined	Final formulation has been defined. Equipment train for production has been chosen	Final packaging process has been defined. Equipment train for production has been chosen
		1.1.2	Value-chain map	Full value-chain map includes yield rates, process times, and all secondaries (for example, solvents)	Full value-chain map includes yield rates, process times, and all secondaries (for example, solvents)	Full value-chain map includes yield rates, process times, and all secondaries
		1.1.3	Difficulty assessment	List of difficulties exists, plan to handle is spelled out (QA, HSE, supplier capabilities, etc)	List of difficulties exists, plan to handle is spelled out (QA, HSE, supplier capabilities, etc)	List of difficulties exists, plan to handle is spelled out (QA, HSE, supplier capabilities, etc)
		1.1.4	Starting material choice	Starting materials including GMP/non-GMP status have been confirmed	Starting materials including GMP/non-GMP status have been confirmed	

### Measuring progress

At the heart of the quality-gate process is a set of metrics that evaluate every aspect of technical development—the process, product, costs, legal, and regulatory status, and CMC status (Exhibit 2). These are defined as “maturity metrics,” their purpose being to measure progress towards developing and manufacturing a low-cost, high-quality product that

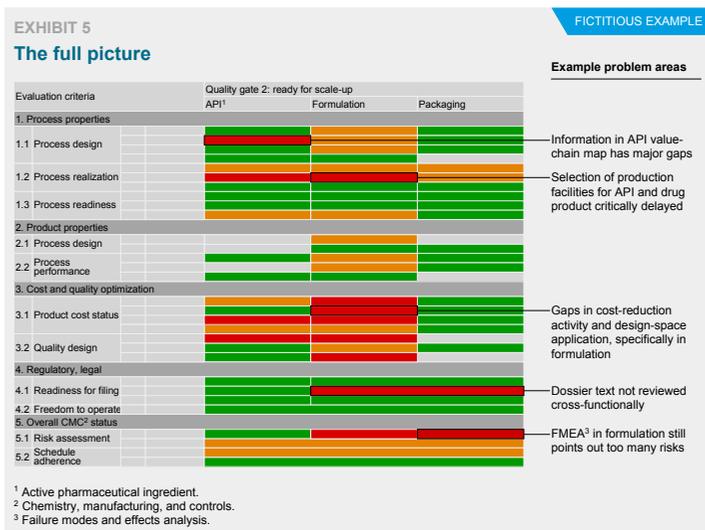
is delivered on time. Each function is responsible for evaluating progress in its own area of expertise. However, given the cross-functional nature of the work, other functions in the quality-gate team need to endorse each evaluation.

A smart mix of qualitative and quantitative metrics can effectively evaluate progress against defined targets without greatly increasing the workload of the teams involved. So, for example, a quantitative measure of process-design maturity, itself a sub-category of process maturity, would be the number of impurities in the formulation. Qualitative metrics are needed for areas that are less easy to define, such as “freedom to operate,” that is, whether a product infringes any patents.

The metrics are the same at each quality gate, though the targets are adjusted to reflect the different phases of development. A traffic light color-coding system then indicates the extent to which the project is on target, again using standardized definitions (Exhibit 3).

This coding system is applied to every sub-category of the evaluation criteria. Exhibit 4 illustrates this for process design, showing the current status of the active pharmaceutical ingredient (API), formulation, and packaging against certain targets. The warning signs here are that there is no map of the API value chain and solvent usage has not been analyzed.

Importantly, the metrics do not only record project progress to date, they



also include an evaluation of the risks involved in taking the project forward to the next stage, and a perspective on how the manufacturing process will evolve towards commercial production standards. Thus, the quality gate encourages agreement about what needs to be done to fix existing problems and to reduce the likelihood of future problems.

The evaluations from all the various project areas are finally combined in a single chart, giving a clear overview of all aspects of the technical-development process, and showing at a glance where risks and problems lie and decisions need to be made. The chart effectively dictates the agenda for the quality-gate meeting (Exhibit 5).

Part of the value of the quality gate is the opportunity it gives participants to learn from the experiences of others on different projects. However, much of the value of the quality-gate process is often realized before the meeting. In the course of collecting the data needed for the quality gate, individual teams often

start addressing problems they uncover straightaway. Individual, preparatory team meetings ahead of a quality gate are therefore important.

Some companies find it helpful to form a dedicated technical-development project office to provide logistical support for the quality-gate process. Ultimately, however, the success of quality gates depends upon those involved adopting a forward-looking, risk-based way of thinking about their work. Participants need to be

comfortable with collecting and sharing data that gives a realistic overview of project progress. And, importantly, companies must raise the profile of technical development so that its value-adding role is recognized.

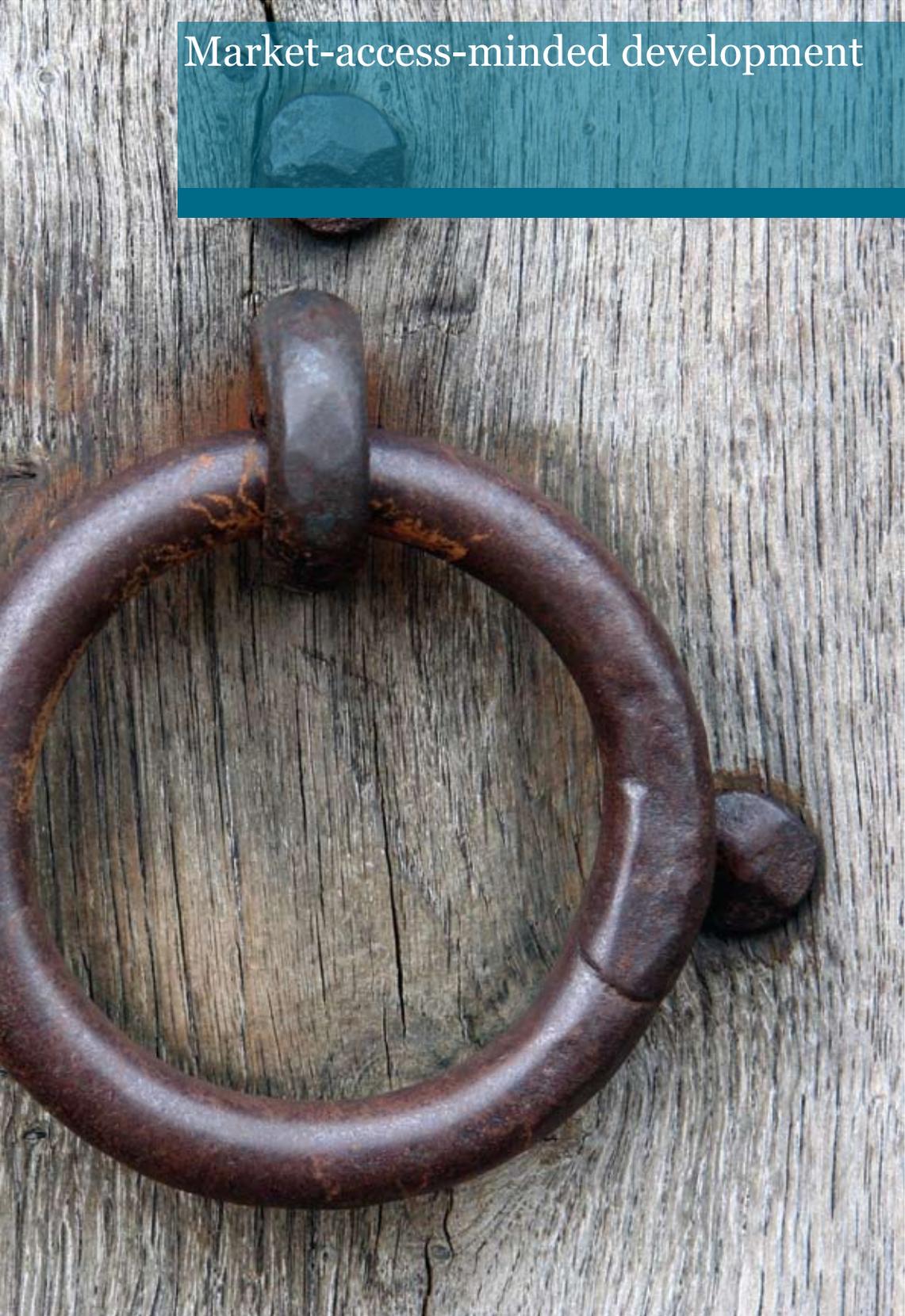
Pilot schemes will help. By using quality gates on one or two key projects with extensive management and administrative support, companies can fine-tune the process and successfully demonstrate the value of the new way of working.

Michele Holcomb ([michele\\_holcomb@mckinsey.com](mailto:michele_holcomb@mckinsey.com)) is a principal in McKinsey's New Jersey office. Martin Møller ([martin\\_moller@mckinsey.com](mailto:martin_moller@mckinsey.com)) is a principal in the Copenhagen office. Paul Rutten ([paul\\_rutten@mckinsey.com](mailto:paul_rutten@mckinsey.com)) is an associate principal in the Amsterdam office.

<sup>1</sup> The process has various other names, including chemistry, manufacturing, and controls (CMC); CMC development; process research; pharm sci; and chemical pharmaceutical development.

<sup>2</sup> Ted Fuhr, Michele Holcomb, and Paul Rutten, “Why ‘Quality by Design’ should be on the pharma CEO’s agenda,” <http://operations-extranet.mckinsey.com>.

<sup>3</sup> Anil G. D’souza, David J. Keeling, and Richard D. Phillips, “Improving quality in pharma manufacturing,” [mckinseyquarterly.com](http://mckinseyquarterly.com), September 2007.



## Market-access-minded development

**A new way of conceiving clinical development will uncover the drugs bound for commercial success early on. Think value.**

Matthias Evers

Recent research shows that clinical differentiation against the standard of care is the best indicator of a drug's future commercial success.<sup>1</sup> This finding is borne out by the high-profile failures of products identified as being insufficiently differentiated or as cost-inefficient by institutions such as Germany's Institute for Quality and Efficiency in Health Care and the UK's National Institute for Clinical Excellence, and by managed-care organizations. To win market acceptance as well as regulatory approval, pharmaceutical companies must develop drugs that are differentiated and conduct trials for them in a way that demonstrates their value convincingly.

We have coined the term "market-access-minded development" to describe a new way of crafting clinical strategies and designing trials. The crucial difference is that, as early as the beginning of Phase II, clinical teams start thinking beyond the endpoints, patient segments, and study designs that line the path to approval, and focus on value as well. Not all stakeholders will define value in quite the same way, but most seek evidence of better patient outcomes.

### **Characteristics of this new approach to development**

Market-access-minded clinical development is likely to include some or all of the following four characteristics.

**It will be more value-focused and forward looking.** Delivering a specified efficacy after four weeks of treatment might be the right endpoint to choose as far as winning approval for a drug is concerned. But if the specification does not differentiate the drug against the standard of care at the time of launch, it will not help to make it successful commercially. Clinicians and their teams thus need to move beyond the "target product profile" to think through how the unmet needs of patients, physicians, and other stakeholders can be met, and to design trials that reveal what added value the drug can deliver. It is therefore important not to tweak the profile to suit the clinical data, but regularly to check that it still reflects stakeholders' needs. Because these needs evolve, and are likely to differ by region and perhaps country, a clear definition of desired patient outcomes and cost effectiveness needs to be part of the traditional approach to building a target product profile.

**It will be more targeted and front-loaded.** Clinical teams will apply patient outcomes, cost-efficiency and economic milestones early in development, that is, before they initiate more expensive late-stage Phase III trials. Just as "key opinion leaders" from science are currently involved early in development, so key opinion leaders from the realm of health economics (such as health economists

from academic institutions) will be brought in to assess and communicate the cost effectiveness of a new drug. Their expertise, as well as the clinical team's creativity, will deliver innovative clinical strategies that characterize a drug's potential value to patients during proof-of-concept trials and in Phase II.

**It will be more rigorous.** Instead of relying on a tiny group of senior experts who “just know” from experience what trial designs to pursue, market-access-minded clinical development is more rigorous and systematic. It is structured in four steps: define which of a drug's claims are the most valuable to prove, identify the most suitable patients, define the most relevant endpoints, and deploy the most cost-efficient design.

**It will be more cross-functional.** Clinical strategy and trial design are often considered the home turf of the senior brand clinician. In the new world, clinical strategy and trial design will become a more cross-functional effort involving a broader group of experts: pricing and reimbursement specialists to provide the payor's point of view, health economics and outcomes researchers to advise on the endpoints that demonstrate cost effectiveness, marketing experts to give the customer's perspective, modeling and simulation specialists and biostatisticians to indicate the most suitable patients and cost-efficient designs. The composition, *modus operandi*, and performance management of the clinical-project team as well as individual job profiles will increasingly need to guarantee such cross-functionality.

In summary, market-access-minded development is value-driven. Right from the outset, when clinical trials are set up to demonstrate proof of concept,

development and clinical teams must think beyond the underlying scientific mechanisms and consider how to generate early insights into the way the compound might address patient needs. Armed with such insights, the clinical strategy can be crafted to focus on value in a single indication and across indications. Execution plans for the likely more extensive Phase II trial programs will also focus on value, while simultaneously maximizing speed and mitigating risks. Though this might lead to higher costs and attrition in Phase II, it should facilitate the demonstration of better value in Phase III, yielding data that can be used immediately to negotiate and achieve market access.

### Obstacles which tend to get in the way

It is challenging to transform the way pharmaceutical companies develop drugs, shifting their focus away from achieving approvals to creating value and marketable labels. One impediment is the scarcity of people who excel at clinical strategy and trial design. Even leading companies rely on a very small group of experts whose skills are based on experience and pattern recognition. So far, their abilities have not been codified into a rigorous approach that would support less experienced co-workers. There is little or no training to turn potential into proficiency, and too much reliance on the copying and pasting of supposedly proven designs instead of learning from mistakes and innovating.

Other barriers are cultural and organizational. They include an ingrained habit of focusing on clinical parameters rather than value, risk aversion (it is easier to replicate trials with endpoints that have led to registration in the past than to work

on new designs), functional silos, and a poor interface between development and commercial. Clinicians' reluctance to accept that cross-functional efforts are now essential, and a shortage of capabilities in emerging disciplines such as health economics, outcomes research, and the gathering of payors' views, are further inhibiting factors.

There are no instant solutions: development leaders need to think systematically about all the functions and departments at the interface between development and commercial, and ask

what needs to change in order to shift their company's approach from a focus on approval to a focus on patient and broader stakeholder value. Cultural change is likely to be necessary, as are new capabilities.

Convincing in-house success stories help to break the mould. We urge managers to start by conducting pilots with selected (early) project teams. When high-profile teams embrace market access as an additional challenge and demonstrate the benefits of the new approach, it becomes easier to initiate more fundamental and widespread change.

Matthias Evers ([matthias\\_evers@mckinsey.com](mailto:matthias_evers@mckinsey.com)) is a principal in McKinsey's Hamburg office.

The author would like to thank Valentina Sartori, Jerrina Eteen, and all other colleagues involved in McKinsey's Value-Driven Drug-Development work.

<sup>1</sup> Maria Gordian, Navjot Singh, and Rodney Zimmel, “Why products fail in Phase III,” *In Vivo*, April 2006.



## Changing with the times: how pharmaceutical companies should respond to regulatory change

Drug companies need to make organizational changes to meet the latest patient safety regulations in the United States. But as the regulations evolve, companies will also need to work openly with all stakeholders to ensure the right outcome.

Maria Gordian and Richa Pande

In response to concerns about patient safety, pharmaceutical risk-management programs in the United States have become increasingly stringent. Since March 2008, the Food and Drug Administration (FDA) has required manufacturers to submit a risk evaluation and mitigation strategy (REMS) to manage known or potential safety risks related to certain drugs based on available safety data,<sup>1</sup> and thus to help ensure a drug's benefits outweigh its risks. As a result, pharmaceutical companies need to make changes across their organizations, paying more attention to safety at each stage of drug development and building new organizational capabilities.

But the regulatory environment is still evolving. Although the FDA has established a broad framework for risk management, many details have yet to be settled. It is still unclear, for example, how the level of risk is decided. And larger issues, such as how best to measure the impact of a REMS, have not been addressed. Moreover, while patient safety remains the paramount goal, it is important not to overburden the health care system, or to prevent some patients from receiving potentially life-saving drugs.

Input from all stakeholders—pharmaceutical companies, pharmacists, health care providers, and patients—is

needed to reach the right regulatory outcome. Here, pharmaceutical companies have a key role to play. Managing as they do the conception, development, testing, and launch of new treatments, they have access to data that will be invaluable in efforts to improve risk management, helping all stakeholders to make objective decisions about the safety of medicines and the risk/benefit trade-offs.

### A brief history of pharmaceutical risk management

Regulatory risk-management programs are by no means new. However, the trend has been towards broader and tighter regulation.

In 2004 and 2005, more than 20 drugs were withdrawn from the US and European markets following a wave of injuries and deaths that were blamed on new prescription drugs. Many lawsuits ensued, and public officials, including legislators and the FDA, came under intense public pressure to place more emphasis on drug safety.

Prior to 2005, risk-management programs were implemented on an *ad hoc* basis. Mitigation consisted largely of education programs and changes to product

labeling, although occasionally distribution was restricted too. Clozaril's "no blood, no drug" program in 1990, for example, required schizophrenia patients to register for blood tests to monitor the risk of leukopenia and agranulocytosis.

When studies revealed that the existing risk-mitigation programs were not effective, the FDA took steps to standardize and improve them further, issuing its guidance on risk-minimization action plans (RiskMAPs) in March 2005. RiskMAPs consisted of three elements: targeted education and outreach programs; reminder systems, which were essentially more elaborate education programs; and performance-linked access systems or restrictive-distribution programs.

Safety concerns and public and political pressure continued to mount, however, and the FDA now imposes REMS requirements on those drugs it feels warrant tighter safety assessment and management. According to one of its Senate sponsors, the REMS legislation is intended "to ensure that drug safety is not an afterthought but an integral part of the process throughout the life of a drug." Exhibit 1 shows the various potential REMS requirements: an approved medication guide; a communication plan for health care providers; elements to assure safe use, such as physician training and certification, and patient education; an implementation plan; and a timetable for assessment of the REMS.

Since implementation, about 25 percent of new molecular entities

(NMEs) and novel biologics, and dozens of marketed products, have received REMS requests from the FDA. Between January and June 2009, however, about 45 percent of the 38 new products approved were required to have post-market commitments, indicating that the proportion of products requiring a REMS is likely to rise.

REMS requirements can have implications for everyone in the pharmaceutical value chain. The fact that certain medicines might be available only through certified physicians and pharmacies could prove problematic for patients living in remote areas or with limited access to the health care system. Pharmacists might have to seek additional certification and provide more patient education, and some might need to comply with risk-management plans aimed at helping to prevent drug abuse and overdose. Physicians who wish to prescribe certain drugs might also need to obtain additional certification, spend more time educating patients, and adhere to restrictions on which patients can receive some treatments. Even the FDA faces challenges, trying as it is to meet its expanded responsibilities with limited resources and budget.

EXHIBIT 1 Potential REMS requirements	
	Description
Medication guide	<ul style="list-style-type: none"> <li>Leaflet for patients to inform them of the risks and benefits of the medication. Distributed by the pharmacy when a prescription is filled</li> </ul>
Communication plan	<ul style="list-style-type: none"> <li>Educational materials for health care providers to support REMS implementation, for example, letters, advertisements in trade journals, memos to professional societies</li> </ul>
Elements to ensure safe use	<ul style="list-style-type: none"> <li>Active management of safety risks including:               <ul style="list-style-type: none"> <li>Prescriber education and certification</li> <li>Dispenser certification</li> <li>Restricted administration (for example, given only in hospitals)</li> <li>Dispensation to patients only with evidence of safe-use conditions (for example, laboratory test results)</li> <li>Patient monitoring</li> <li>Patient registry</li> </ul> </li> </ul>
Implementation system	<ul style="list-style-type: none"> <li>System to monitor, evaluate, and improve implementation of "Elements to ensure safe use"</li> </ul>
Assessments	<ul style="list-style-type: none"> <li>Reports to the FDA on the implementation and effectiveness of REMS at least 18 months, 3 years, and 7 years post-REMS approval</li> </ul>

For their part, pharmaceutical companies must now develop more sophisticated programs to communicate with all those who prescribe, dispense, and administer a drug with REMS requirements, and manage and monitor them too. But their approach to risk management will have to extend far beyond this.

### Organizational change

To operate efficiently and effectively in the new REMS environment, pharmaceutical companies will need to adopt new decision-making and management processes, adjust their organizational structures, and develop new capabilities for managing risk from early development through to post-loss of exclusivity (LOE). Specifically, they should:

#### Anticipate and identify possible safety risks, and communicate early with the FDA

Companies should be able to forecast if their drugs will be subject to REMS requirements—there should be no surprises. Identifying potential safety risks early in development must become a habit, as must managing those risks. If a drug seems to be causing elevated liver enzyme levels in early development, for example, every effort should be made to collect the data that explains the problem and to share it with the regulator, not only to demonstrate adequate diligence but also to address properly any concerns about the potential frequency and severity of safety events. By sharing data with regulators earlier and more frequently during development, companies will be able to help shape the understanding of a drug's risks and benefits, and whether a REMS might be required. They might even find that the targeted clinical profile of the drug changes so that it can be used in certain circumstances where it

is recognized that the risks outweigh the benefits. Furthermore, companies will be able to avoid situations whereby the FDA imposes new trial requirements late in the development cycle, creating unforeseen delays and costs.

#### Build efficient REMS-management processes

Companies need to establish efficient management processes not only to meet FDA requirements promptly—they have 30 days to produce a medication guide, for example—but also to prevent any costly launch delays. Although some requirements can be satisfied only on a case-by-case basis—new clinical trial requirements, post-marketing studies, or surveillance programs, for example—others, such as the writing of medication guides, can be met by standardized processes that will speed compliance, lower expenses, and improve safety.

#### Incorporate safety, risk-evaluation, and risk-mitigation considerations at every stage of a product's life cycle

While the FDA typically imposes REMS requirements when a product is approved or when new safety data becomes available for a product already on the market, risk evaluation and mitigation should inform key decisions throughout the life cycle of a drug. Actively addressing safety concerns raised by any preclinical or clinical data should be a priority, and key development decisions (including go/no-go choices, safety and efficacy endpoints, and clinical trial design) should all be considered from a REMS perspective. A company's medical, science, regulatory, legal, and commercial teams should all be involved. Exhibit 2 describes the triggers that should prompt team action. In early development, for example, knowing that a mechanism of action has safety issues or that other products in the same class have

REMS requirements should trigger the collection of more data. It should also prompt discussions with regulators as early as Phase II or III to make sure there is agreement on what data might be required to establish the product's safety profile. Similarly, new safety signals as a result of pharmacovigilance or post-marketing studies might prompt a REMS requirement at extremely short notice, so companies need to be prepared to meet tight deadlines and avoid civil penalties.

#### Understand the financial implications

A REMS will affect budgeting decisions and commercial value. Companies therefore need to factor it into their revenue forecasting, product-development investments, and the costs related to product launch, sales, marketing, and post-marketing. REMS-related costs and revenue implications also need to be considered when valuing assets for portfolio prioritization, and for licensing and acquisition purposes.

#### Develop new capabilities

Most companies will need to acquire new capabilities so that they can respond quickly to and manage REMS requirements, mine internal and external data more thoroughly for earlier warning signs about risks, and continuously monitor and assess risk-management programs both for reporting purposes and to improve them. While the delays and costs associated with these efforts might be particularly onerous for smaller companies, every company will

need the capabilities required for risk management, from product development to commercialization.

#### Regulatory work in progress

The new REMS regime is still work in progress. Operational aspects are being worked out as the FDA and industry together decide the best way to meet REMS requirements. Companies are currently required to lay out detailed plans, which the FDA will then approve. But with time and experience, it is likely that there will be greater clarity and guidance on what constitutes best practice.

For the time being, there are many unanswered questions. How, for example, can the success of a REMS be measured? Is it feasible to collect and analyze data in a way that will reveal what works and what does not and to modify programs accordingly? And how should "success" be defined if restricting access to certain medicines for safety reasons denies life-saving treatments to patients prepared to make a different cost/benefit trade-off?

Thought also needs to be given to how safety can be improved without overburdening the health care infrastructure. The FDA is sensitive to practitioners' concerns about having to comply with different companies' REMS for drugs in the same class, and so has been considering whether a single, class-wide REMS might be appropriate for extended-release opioids. To help it in its deliberations, it has called for a series of meetings with different stakeholders to discuss the possible implications, including how competitors might work together to develop a common REMS; the IT infrastructure required to support a REMS that would cover more than 20 million prescriptions a year; how its effectiveness would be measured; and who should be responsible for operating the REMS program.

It is in these kinds of communications between all stakeholders that pharmaceutical companies have an important role to play. With so much data at their fingertips, they are an essential participant in discussions aimed at improving the safety of medicines, understanding the risk/benefit trade-offs, and managing those trade-offs. Pharmaceutical companies can and should take the initiative whenever possible to facilitate the open exchange of information, actively participating in debate and, when appropriate, themselves bringing together stakeholders by convening round-table or working sessions. The aim is to establish a transparent process whereby all stakeholders work towards shaping the right regulatory outcome—an effective approach to ensuring patient safety.

Maria Gordian ([maria\\_gordian@mckinsey.com](mailto:maria_gordian@mckinsey.com)) is a principal in McKinsey's New York office and Richa Pande ([richa\\_pande@mckinsey.com](mailto:richa_pande@mckinsey.com)) is a consultant in the New Jersey office.

**EXHIBIT 2**  
**Safety triggers**

Timing/phase of product development	Early development Discovery and preclinical testing Clinical development (Phase I)	Late development Clinical development (Phases II and III) NDA <sup>1</sup> submission NDA review	In-market Marketed/mature products requiring REMS (reactive versus proactive)	Post-LOE Products that have lost market exclusivity
<b>Example triggers</b>	<ul style="list-style-type: none"> <li>Mechanism of action with known safety issues</li> <li>Class labeling/REMS</li> <li>Safety issues identified through preclinical, toxicity, and Phase I data</li> </ul>	<ul style="list-style-type: none"> <li>FDA requirement for product in response to data/submission</li> <li>Class labeling/REMS</li> <li>Safety issues identified through Phase II or Phase III clinical data</li> </ul>	<ul style="list-style-type: none"> <li>FDA requirement for product in response to data/submission</li> <li>Class labeling/REMS</li> <li>New safety signals</li> </ul>	<ul style="list-style-type: none"> <li>FDA requirement for product in response to data/submission</li> <li>Class labeling/REMS</li> <li>New safety signals</li> </ul>

<sup>1</sup> New drug application.

<sup>1</sup> Their own data or that available for similar drugs.

## The anatomy of attrition

Research tracking the attrition rates of more than 3,000 compounds in each phase of development reveals recent drug failure trends.

Navjot Singh, Rodney Zimmel, Maria Gordian, and Áine Denari

More than 90 percent of compounds that enter Phase I trials are destined to fall out of the development pipeline, making drug failure a headline issue for the pharmaceutical industry, academics, analysts, and drug regulatory authorities. Attrition rates of this magnitude are the single biggest cause of the industry's R&D productivity challenge.

Many companies have made concerted efforts to cut attrition rates (see sidebar, "Cutting the losses"), and much research has been conducted to explain why it continues to be so high. In 2006, for example, McKinsey research showed that almost half of all failures in Phase III were due to a lack of demonstrable efficacy of a drug compared with a placebo.

To update our research, we conducted an outside-in analysis of pharmaceutical R&D attrition rates between 1996 and 2007. Using publicly available data, we tracked each phase of more than 3,000 compounds in development (see sidebar, "Methodology"). By tracking phases rather than full projects, we were able to capture what has been happening quite recently in the industry, and so assess trends in more detail. Here are the main findings.

### Overall clinical success rates have worsened in the past 12 years

Across the industry, the success rate of drugs fell between 1996 and 2007, and worsened in the second half of that period. Almost all major biopharmaceutical companies followed the trend (Exhibits 1 and 2).

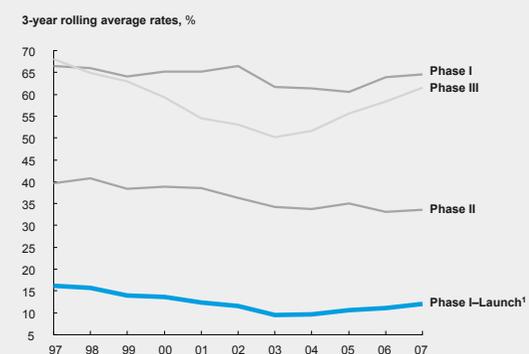
### Higher attrition in Phase II is the main reason for declining success

As the exhibit shows, attrition during Phase I has been relatively stable. After a dip earlier in the analysis period, success rates in Phase III are now recovering, leaving attrition in Phase II as the single biggest reason for declining success. Outside-in analysis suggests a 16

EXHIBIT 1

#### Overall decline

Across the industry, the success rate of drugs has fallen, mainly because of the rising level of attrition in Phase II



<sup>1</sup> Phase I-Launch success rate calculated as (Phase I 3-year rolling average x Phase II 3-year rolling average x Phase III 3-year rolling average x 90%); using 90% assumption for regulatory/launch success.  
Source: McKinsey R&D Attrition Database (excluding formulations); Pharmaprojects; McKinsey analysis

## Cutting the losses

The approaches that companies are using to reduce attrition rates fall into four main categories:

### A better understanding of the causes of failure

Companies have invested in developing internal databases that track in great detail the history of compounds, pinpointing when they failed and why. These databases have proved effective in helping companies understand the root causes of failure and to take corrective action—for example, by changing milestone-driven guidelines and introducing new assays. This input also provides useful feedback to discovery scientists who design new molecules.

### Earlier identification of drugs likely to fail

Scientists in the biopharmaceutical industry understand that attrition is part of development. The primary strategy of most biopharmaceutical companies has therefore been to identify at an early stage those drugs that are likely to fail, and to stop their development and avoid wasting money. There are three ways companies are pursuing this.

First, companies are investing in more small-scale in vivo and in vitro testing in late discovery to screen out compounds, especially those that might have toxicity issues. By contrast, efficacy biomarkers have been moderately successful because of difficulties in human predictability.

Second, companies are increasingly pursuing high-throughput, early clinical-screening strategies to enable faster and more efficient clinical screening to test for efficacy in patients. These techniques include adaptive clinical-trial designs and early patient testing, and the building of infrastructure to enable faster patient screening.

Third, companies have enhanced their governance models and team structures. Over the years, governance has become the focus of efforts to improve the survival rates of candidate compounds. Companies have improved their decision-making processes, governance philosophies, governance architecture, and the tactics and operations of governance.

percent decrease in Phase II survival since 1997, leading to a survival rate of 33 percent in 2007. Internal analysis of some companies suggests a Phase II survival rate of between 15 and 20 percent, particularly when survival rates are assessed by indications (although indication attrition is not always fully reported publicly).

### Causes of Phase III failures are shifting

Our previous research into drug failures in Phase III<sup>1</sup> identified lack of efficacy compared with a placebo as the main reason. This continues to be the single biggest cause of failure across the

industry, as illustrated in Exhibit 3, and remains a significant opportunity in terms of reducing attrition rates—failures due to lack of efficacy in Phase III should be negligible if Phase I and Phase II trials are conducted well.

However, further disaggregation of failures across five-year periods reveals three more interesting findings (Exhibit 4). First, the number of failures due to lack of efficacy compared with a placebo is falling. Second, despite a number of high-profile Phase III failures owing to safety concerns, the relative fraction of such failures as a percentage of the

### Finding alternative uses for drugs

Ultimately, of course, launching a drug is more valuable than stopping one that is likely to fail. It is vital therefore to explore alternative indications for a candidate drug that fails in the primary indication, as long as the exploration is justified. Many companies have tried to systematize searches for alternative indications. Some, for example, have created groups that look for common pathways in the discovery phase, others focus on clinical investigations.

### Modifying portfolio strategies

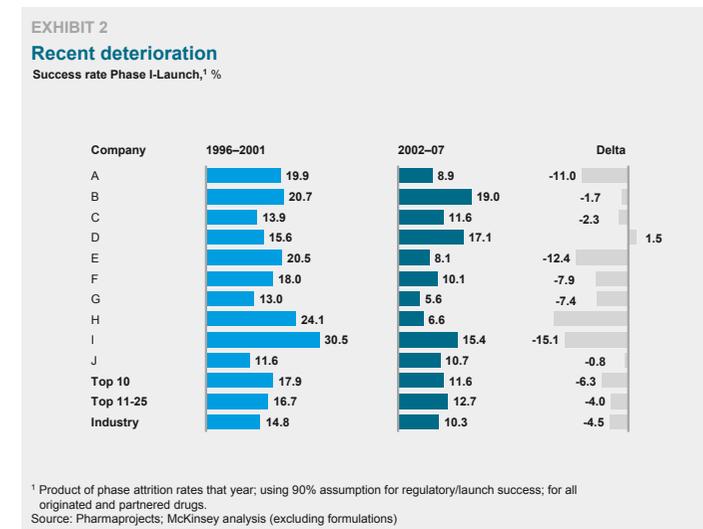
The choices that companies make in relation to their portfolios are one of the most important ways of reducing attrition. There are four that companies typically consider.

The balance of proven and unproven mechanisms in the portfolio. Even though the expansion of target space in the clinic has been minimal, the attrition rate of unproven mechanisms is almost twice that of proven mechanisms. What ultimately matters is meaningful clinical differentiation. How best to balance proven and unproven mechanisms in the portfolio remains a much-debated topic, and different companies have arrived at different conclusions.

The number and diversity of back-up candidates. Advances in high-throughput screening have made it easier to identify hits, but decisions about the diversity of hits, and the timing of the back-up, are critical and can have significant impact on overall program survival.

How to balance spending on life-cycle management and label expansion against investments in new molecular entities. Whilst the exact balance will differ by company, many have increasingly focused on life-cycle management and label expansion in response to increased regulatory stringency.

Whether to move away from the industrialized model for R&D. This model linked a set annual number of drug candidates with incentives and bonuses. The alternatives are a new model of proof of concept, or a model that culls the number of preclinical candidates in the hope that only those with the best chances of success will be advanced.

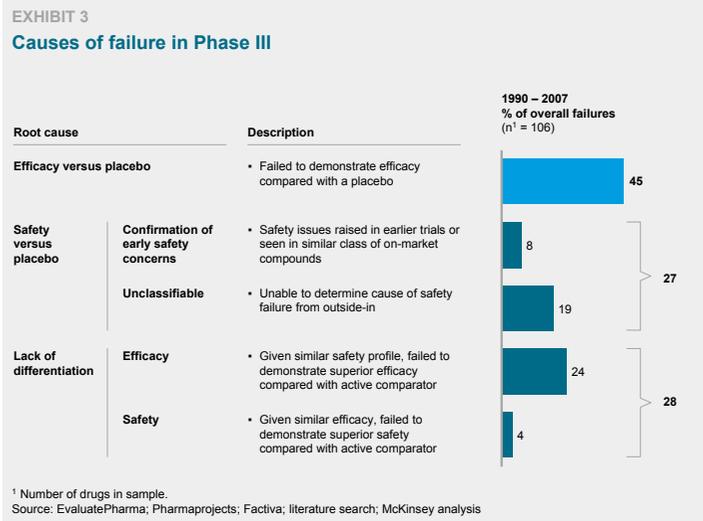


total has remained stable. However, it is worth noting that the number of failures in Phase III arising from safety concerns that confirm doubts raised earlier in development has increased. This could imply that the industry is getting better at spotting potential risks in earlier phases, but that it fails fully to evaluate the safety-risk benefits, leading to more expensive failures later in development.

Third, the number of failures due to lack of commercial differentiation has risen slightly during the past five years—an increase probably driven by the higher demands of payors and access agencies. However, the question remains whether these failures could have been avoided by Phase II trials that better reflected a clear understanding of customer needs.

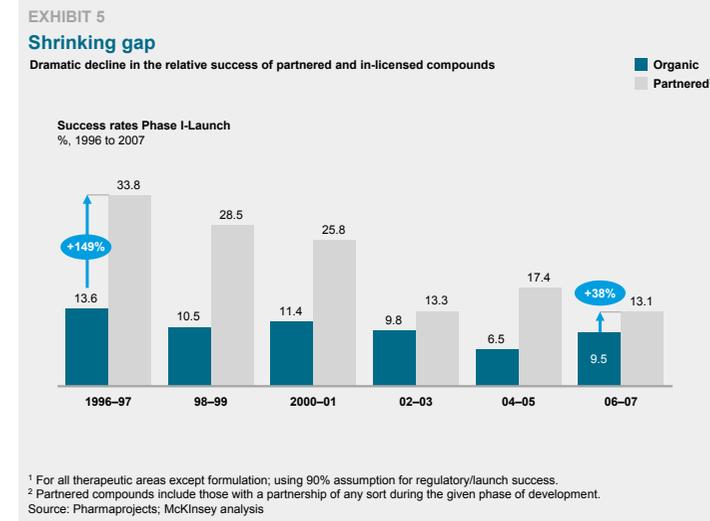
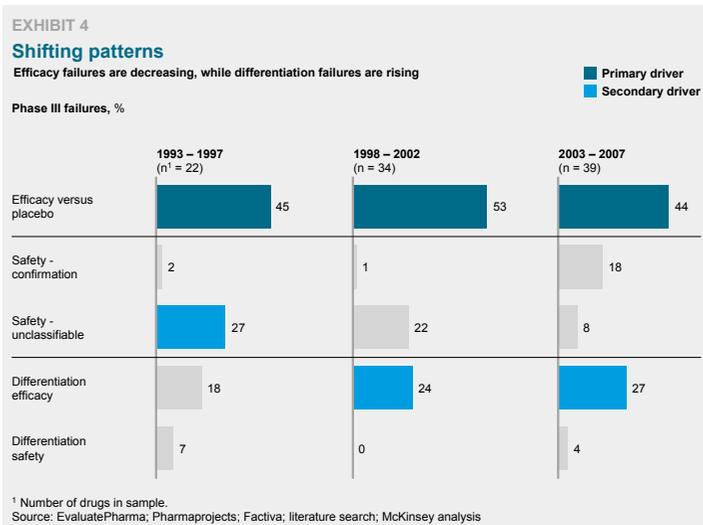
**Success rates have fallen in most therapeutic areas**

Although the sample size is relatively small because of the limitations on establishing root causes of failures from the outside-in, it would appear that success rates in most therapeutic areas (TAs) have declined. The average decline is 4.5 percent (Exhibit 4), but the figure varies by TA. For example, oncology success rates have dropped by 3.2 percent and endocrine by 13 percent. Only cardiovascular success rates have risen, and these by just 3 percent.



**A dramatic decline in the relative success of partnered and in-licensed compounds**

Compounds developed in partnerships were once the stars of drug development, far exceeding the success rates of those originated internally (Exhibit 5). Although partnered compounds still outperform “organic” compounds, the gap has shrunk. Now, the success rate of partnered projects is 38 percent higher than that of organic ones, compared



with almost 150 percent higher 12 years ago. The biggest pharmaceutical companies seem to achieve similar results in both categories—a quite different story from 12 years ago, when their partnered compounds had a 72 percent higher success rate than their organic compounds. The shrinking gap no doubt partly reflects the efforts made by many companies to manage their internal portfolios more rigorously.

## Methodology

*We conducted an outside-in analysis of pharmaceutical R&D attrition rates in the past 12 years, tracking the phases of more than 3,000 compounds in development between 1996 and 2007. Our research concentrated on orally ingested, small-molecule drugs (excluding biologics such as vaccines) produced by large pharmaceutical companies, using Pharmaprojects data rather than companies' self-reported data.*

*We defined a drug as having "failed" if the trial ended early (but excluding those that ended early because of strongly positive results), or if the trial failed to produce the results that would ensure drug approval. We analyzed the cause of failure for those drugs for which there was sufficient public information available, considering their efficacy and safety (as compared with placebos), and comparing them with similar drugs already on the market. Our Phase III analysis consisted of outside-in investigations to assess the causes of failure: analyst reports, press releases, and other sources of public information. A panel of clinical experts reviewed our material to ensure we had interpreted the data correctly. Though this data set is limited, we believe it to be more robust than the internal information on trial failures available at any single pharmaceutical company.*

*The analysis only begins tracking phase success of a compound after the compound enters Pharmaprojects, rather than including it as a success in earlier phases. This avoids over-counting the number of successes and undercounting the number of failures in earlier phases. Each phase of each project was tracked and analyzed independently based on the year it ended that particular phase (the phase "exit year"). By tracking phases rather than full projects, we were able to look at composite attrition numbers for the industry that reflect what has been happening quite recently. This allows for a more detailed assessment of trends over time than can be achieved by other analytical approaches, which tend to be limited in sample size.*

*The analysis highlights trends in R&D attrition rates, but also looks at the impact of other factors such as therapeutic area (TA), company size, technology, and licensing status.*

*Historical in nature, the analysis makes no attempt to predict future trends, which are driven by factors not easily modeled on the basis of historic data, including regulatory changes, portfolio decisions, market access issues, and scientific evolution. Benchmarking at the TA level within individual companies (or even entire small companies) is not always meaningful because of low sample size.*

Navjot Singh ([navjot\\_singh@mckinsey.com](mailto:navjot_singh@mckinsey.com)) and Maria Gordian ([maria\\_gordian@mckinsey.com](mailto:maria_gordian@mckinsey.com)) are principals in McKinsey's New York office, where Rodney Zimmel ([rodney\\_zimmel@mckinsey.com](mailto:rodney_zimmel@mckinsey.com)) is a director. Áine Denari ([aine\\_denari@mckinsey.com](mailto:aine_denari@mckinsey.com)) is a consultant in McKinsey's Dublin office.

The authors would like to thank Carla Zwaanstra, Susan Waters, Kendra Buzzell, Kenneth Yoon, C.J. Pavia, Kristoffer Famm, and Tony Tramontin for their contributions.

<sup>1</sup> Maria Gordian, Navjot Singh, and Rodney Zimmel, "Why products fail in Phase III," *In Vivo*, April 2006.

## What gives drugs the X factor?

Despite the long development process, more than two-thirds of new drugs launched on the market are commercial failures. Recent research uncovers what distinguishes those that succeed.

Ramesh Hariharan and Navjot Singh

Few industries have a more complex value chain than the pharmaceutical sector. From the first stage of discovery to commercial launch, the R&D process is long (from 11 to 14 years per drug, on average) and expensive (from half a billion to a billion dollars per drug). To cover the costs, pharmaceutical manufacturers traditionally rely on a few blockbuster compounds to make up for the underperformers.

But the blockbusters are few and far between. Research has shown that only about 30 percent of compounds entering the market recover their risk-adjusted cost of R&D.<sup>1</sup>

Why do so many branded drugs fail commercially? And what distinguishes the few that succeed? Companies have various theories. Some will point to the level of unmet needs being addressed, others to competition from branded drugs and generics. Many believe that the novelty of a compound's mechanism of action (MOA) is the critical factor. Yet few in the industry have studied the question explicitly. Although the 70 percent failure rate at this stage is comparable to that of Phase II clinical trials, post-launch failures receive far less attention. A 2004 study showed that the novelty of a compound does not correlate with success in the market, but it did not go on to examine other factors.<sup>2</sup>

To understand better the causes of commercial failure, we analyzed 50 new molecular entities launched between 1997 and 2000. We considered six possible factors that might influence their success, and found that only two are important: differentiation relative to the existing standard of care, and market size.<sup>3</sup> Of the two, differentiation is the more significant.

The findings have important implications for a wide range of R&D, commercial, and business-development activities, both for compounds in development and for those being launched. Post-launch failures are bound to occur sometimes. But by understanding what drives commercial success and so directing resources more effectively, the industry as a whole—and certainly individual companies—can, we believe, achieve commercial success more than 50 percent of the time.

### Understanding commercial success

Pharmaceutical executives need a better understanding of the variables that determine the success or failure of product launches. We analyzed six potential drivers of success.

1. Differentiation with respect to the standard of care in three areas: safety, efficacy, and convenience

- Novelty of the drug's MOA, that is, a binary assessment that depends on whether a commercial product with the same MOA exists at the time of launch
- Market size at time of launch, measured as the sum of all sales of drugs for a given indication, along with an estimate of future growth<sup>4</sup>
- Competitive intensity at time of launch, including the number of patented competitors, the number of competitors in the pipeline, and the market share of generics
- Unmet medical or market need, assessed in the same three areas as differentiation: safety, efficacy, and convenience
- Period of exclusivity, or the amount of time until generic competitors can enter the market.

We analyzed these variables for 50 compounds released between 1997 and 2000 (see sidebar, "Methodology"). We then performed a regression analysis to identify the most important ones.

Two variables—differentiation and market size—predicted the success or failure of all the compounds examined. Differentiation alone predicted the fate of 77 percent of the compounds, while market size accounted for the remaining 23 percent. The other variables were statistically insignificant.

As shown in Exhibit 1, all ten of the drugs judged to be highly differentiated that were launched in moderate to large markets were successful—that is, they were among the 25 with the highest net present value of the compounds we studied.

All the successful drugs reached annual sales of \$550 million, and most topped \$1 billion. Even in small markets, ten of the 14 differentiated compounds were successful; the only exceptions were four compounds that, despite being targeted at a large market, ended up with narrow labels restricting them to niche segments.

All 11 products that lagged behind the standard of care in any way were failures, regardless of the market size. One drug was approved and launched for the treatment of serious bacterial infections and exhibited life-saving potential for some patients who did not respond to the standard of care at the time of launch. Yet the compound failed because of a poor safety profile (severe cases of inflammation and pain). Similarly, a hyperlipidemia drug failed because it displayed lower efficacy than Lipitor, even though it was safer by some measures (including lower risk of LFT elevation). Another product was launched with the goal of creating a market for a drug that expedited flu recovery. The drug was effective—the average recovery period fell from five days to three—and even offered potential for off-label use for flu prevention

during the flu season. However, it failed to meet the standard for convenience. Two chemicals, the active ingredient and lactose, needed to be placed in an inhalation device before intake.

Additionally, we found that new entrants need to exceed the standard of care only slightly to achieve success. Products that are highly superior to incumbent compounds in safety, efficacy, or convenience do not have a proportionately higher chance of commercial success.

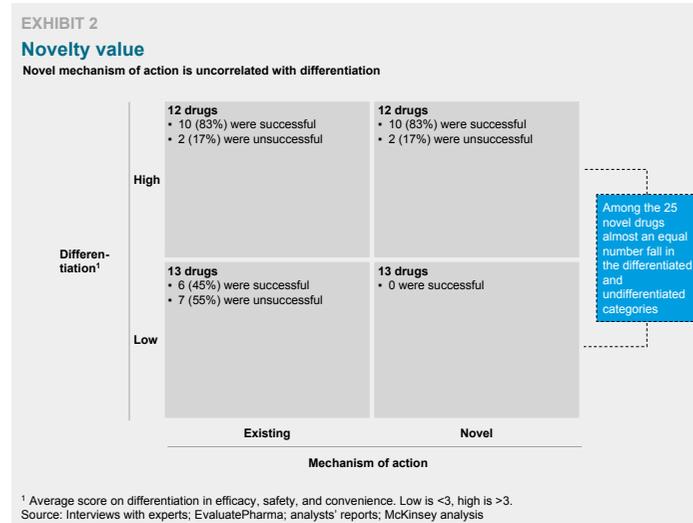
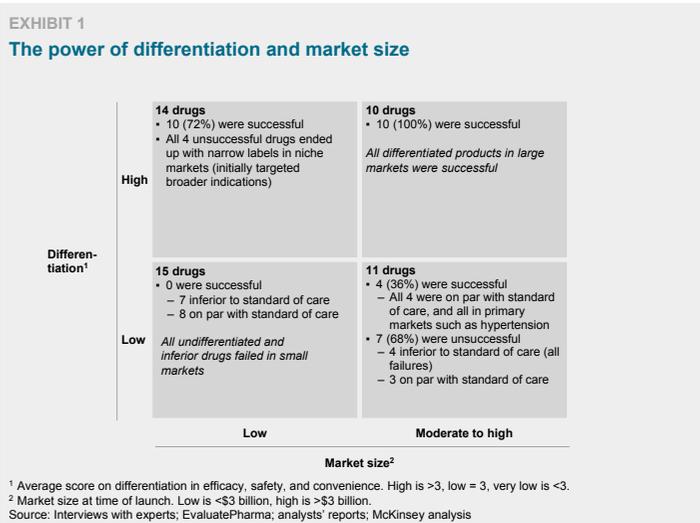
An examination of the undifferentiated products yields a surprise. All successful drugs in our data set of undifferentiated drugs serve large primary-care markets. Why? We believe one reason is that specialist physicians tend to make more discerning recommendations than their primary-care counterparts. Those drugs that simply match the standard can succeed, therefore, provided they address large, primary-care markets and are marketed carefully. In addition, primary-care markets tend to be large enough to allow for well designed promotional campaigns to exploit fine distinctions between different segments of the market.

Why were the other variables insignificant? We can offer rationales for each.

- The number of competitors is less important than how high a bar is set by the standard of care, and hence what the company must do to exceed it.
- We found no correlation between generic market share and success rates of new entrants; differentiated products succeeded even in markets with strong generic share, reflecting payors' willingness to pay a premium for differentiation.
- Unmet needs that are not addressed by the new drug are irrelevant; the needs that are addressed are (at least partially) captured under differentiation.
- Although success rates were greater in markets with higher growth, the difference was statistically insignificant, probably because good brands can take market share from competitors even in stagnant markets.
- While patent duration will correlate to lifetime profitability, it does not affect the success of a launch, which is determined in the first few years.

We found no correlation between differentiation and MOA. We expected

to find that compounds with a novel MOA would be more likely to be differentiated relative to the standard of care. However, among the 25 drugs with a novel MOA, only 12 are differentiated—the same number of differentiated drugs as there are within the group of 25 with a proven MOA (Exhibit 2). More surprisingly, a larger percentage of drugs turned out to be commercially successful in the group with the proven MOA than



in the group with a novel MOA (60 percent versus 40 percent).

### Misplaced energy

Limited understanding of what drives commercial success means energy and resources are misallocated. For example, instead of focusing on defining the current and emerging standard of care, and on outlining a path to differentiation for clinical teams, much effort often goes into sales forecasting—even though the correlation between predicted and actual sales is poor—and the novelty of a compound's MOA.

When developing target product profiles (TPPs), companies often rely strictly on detailed market research, which frequently exaggerates the level that the new product must achieve to be successful. In addition, some companies invest time in developing a lengthy list of variables for their TPPs, even though many of these are unimportant to physicians.

Finally, limited understanding can result in poorly designed clinical trials. For example, development teams often do not incorporate external comparators (especially the drugs that are the standard of care) in Phase II and III trials, fearing that their compounds will not stand up well to them. While this might be wise in some cases, some drugs destined for small, specialized markets would benefit from the comparison in clinical trials.

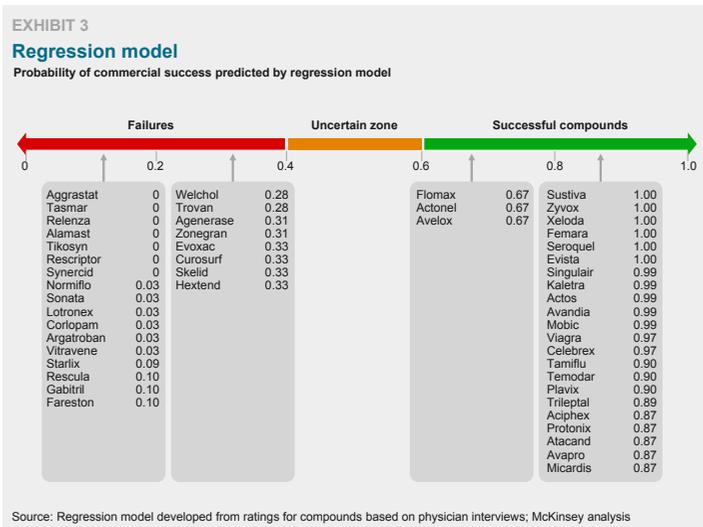
Similar problems crop up in commercial activities. In evaluating in-licensing

candidates and acquisition opportunities, business-development teams often spend a lot of time discussing variables that have little influence on commercial success.

Commercial resources dedicated at the time of launch, and the expectations communicated to shareholders, are often poorly aligned with the chance of commercial success. Even compounds that are known to be inferior to the standard of care (on any one dimension—safety, efficacy, or convenience) get launched as long as they are approved. Sometimes this is understandable given sunk costs: it can pay to launch drugs that are likely to generate positive returns from that point forward. But failing to deliver on expectations communicated at the time of launch is not received well by financial markets.

Moreover, an uncertain understanding of the value of differentiation causes some companies routinely to put off life-cycle management projects in favor of developing new chemical entities.

Finally, where companies do not apprehend the commercial risk, they also



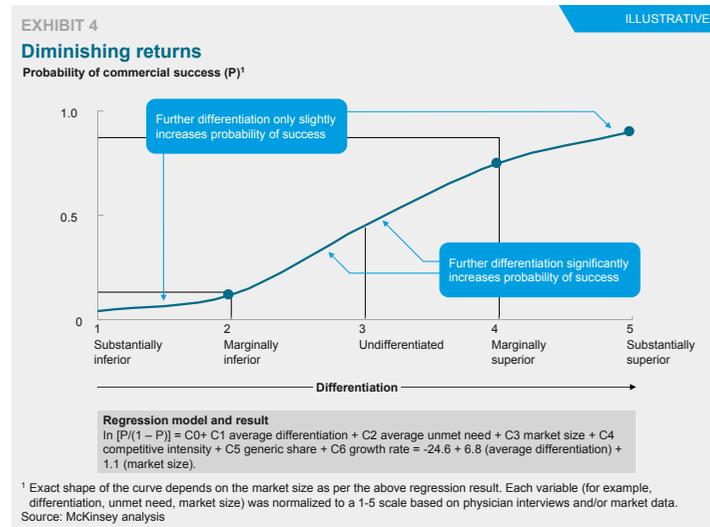
### Methodology

We began our research with all 137 compounds launched from 1997 to 2000. We filtered this list to exclude orphan compounds, non-NCEs (for example, reformulations), and drugs developed by small manufacturers, leaving a sample of 94. Using an NPV analysis, we then identified the 25 top performers (all of which had sales of at least \$550 million; 23 had sales of more than \$1 billion) and the 25 bottom performers (all had sales of less than \$200 million).<sup>5</sup> For those 50 compounds, we rated five potential drivers of commercial success on a one-to-five scale, from “substantially inferior” to “substantially superior.” The ratings were based on two or three expert interviews for each drug and external research. To standardize the quantitative responses from experts, we were explicit in our definitions of each driver. In most cases, the expert evaluations were consistent with external articles. In the few cases where there were disagreements, we supplemented the assessment with additional interviews. Our conclusions are based on the clinical judgement of these experts. We then conducted a regression analysis to assess the relative importance of the potential drivers. We first chose the single variable that helped explain the largest number of successes and failures, then added variables sequentially to show the greatest possible incremental gain in prediction. We stopped adding variables when a sufficient number of compounds were correctly described.

We used logistical regression instead of linear regression, mainly because the former does not permit probabilities to be less than zero or greater than one. The resulting equation from the regression analysis predicts a probability of success, which we then translated to “success” and “failure” depending on the range of value (Exhibit 3). The analysis also demonstrates the law of diminishing returns: that highly differentiated compounds do not have a substantially higher probability of success than those that are only marginally differentiated relative to the existing standard of care (Exhibit 4).

sometimes fail to understand adequately the risk to future cash flows—the enterprise risk. A solid understanding

of the enterprise risk posed by some critical, late-stage assets is an important factor in capital structure decisions, such as the optimal debt-to-equity ratio to maximize shareholder value.



### Implications

Understanding what drives commercial success suggests a clear set of actions for executives.

Companies should emphasize clinical differentiation rather than mechanistic differentiation. This requires simple tools that can help summarize

the key differentiating attributes, and so drive the debate around differentiation. Understanding differentiation should take priority over forecasting, managed-care strategy, pricing, a detailed launch plan, and life-cycle strategy, especially prior to proof of concept.

In addition, for compounds for which efficacy results are available (typically after Phase IIa), an appropriate commercial risk factor should be added to the technical risk of failure to increase transparency into the impact on future cash flows.

In early screenings of several possible in-licensing candidates, business

development could focus more on understanding differentiation relative to the standard of care at the time of launch, and on the market size. (A longer list of attributes is more commonly the focus of extensive discussion.)

Our research should help teams spend their time and effort on the most important activities at the most appropriate junctures during development and product launch. It should thus also help to improve the quality of assessment, the quality of the compounds being launched, and, therefore, the chance of commercial success.

**Ramesh Hariharan** is an alumnus of McKinsey's New Jersey office, and **Navjot Singh** ([navjot\\_singh@mckinsey.com](mailto:navjot_singh@mckinsey.com)) is a principal in the New York office.

We are grateful for the contributions made by Edd Fleming, Mike Pearson, Rodney Zimmel, Michele Holcomb, Laura Furstenthal, Maria Gordian, Roy Berggren, Lara Sullivan, and Philip Ma, all McKinsey colleagues. Padma Sundar deserves a special mention for conducting several of the interviews and the initial analysis.

<sup>1</sup> See Henry Grabowski, John Vernon, and Joseph DiMiasi, "Returns on R&D for 1990s new drug introductions," *PharmacoEconomics*, 2002 Vol. 20, pp.11-29.

<sup>2</sup> See Bruce L. Booth Jr. and Rodney W. Zimmel, "The search for blockbuster drugs," [mckinseyquarterly.com](http://mckinseyquarterly.com), August 2004.

<sup>3</sup> Because our work is based on historical evidence, it assumes the status quo in development, pricing, and payor environments. A radical shift in any of these areas could alter the primary drivers of success for branded pharmaceuticals. The views in this paper are ours alone and are based strictly on market data and interviews with physicians.

<sup>4</sup> We measured market size and growth strictly by sales of all drugs for the indication in question, providing us with an objective measure. We avoided using other possible measures of market size (prevalence, number of patients being treated) as these are harder to assess and are only partial components of a patient flow.

<sup>5</sup> Costs of marketing and sales and R&D used to estimate the NPV are identical to those used by Grabowski et al in "Returns on R&D for 1990s new drug introductions," *PharmacoEconomics*, 2002 Vol. 20, pp.11-29.

## A special kind of success

Mainstream pharmaceutical companies are turning their attention to specialty drugs. Which factors are most important for commercial success or failure?

Mark Gudiksen, Edd Fleming, Laura Furstenthal, and Philip Ma

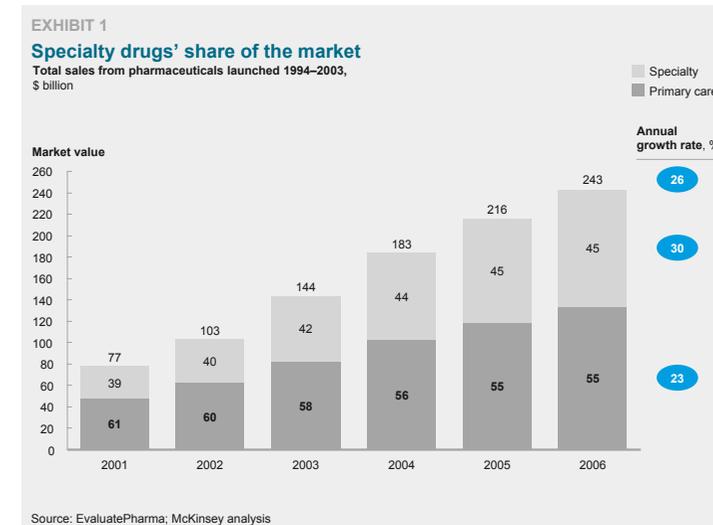
For most of the past 30 years, large pharmaceutical companies have focused mainly on developing and marketing blockbuster drugs that are used by extensive populations of patients. These drugs, such as Lipitor (atorvastatin, Pfizer), Nexium (esomeprazole, AstraZeneca), and Zoloft (sertraline, Pfizer), are usually prescribed by primary-care physicians (PCPs). During the past decade, however, the industry has seen the growing commercial success of specialty pharmaceuticals—drugs prescribed mainly by specialists rather than PCPs. Total annual revenue in the period from 2001 to 2006 of all drugs launched between 1994 and 2003 shows that specialty products are taking an increasing portion of the market,

advancing from 39 percent of total sales in 2001 to 45 percent in 2006 (Exhibit 1).

Although specialty products traditionally have been the preserve of biotechnology pioneers such as Genentech and Amgen, mainstream pharmaceutical companies are increasingly adjusting their focus. The chief executive officers of Pfizer, Bristol-Myers Squibb, and AstraZeneca have all publicly announced their intention to increase their development of specialty and biologic products.

Given the excitement surrounding specialty drugs, we set out to investigate the common factors that contribute to their success. For example, technological

innovations—such as recombinant proteins, monoclonal antibodies, and genome-based diagnostics—tend to make their commercial debuts in specialty markets, indicating a greater focus on novelty in product development. At the same time, modifications of existing drugs, such as new formulations, are quite common in specialty markets given manufacturers' desire to preserve franchises



through lifecycle management and the wish of specialty physicians for improved therapeutic options for their patients. So which factors are the most important for success or failure in this field?

## Basis of analysis

We analyzed all 143 non-generic, specialty pharmaceuticals launched in the United States from 1994 to 2003, using the net present value (NPV) of the first five years of US sales post-launch as our metric of commercial success. We classified drugs as specialty pharmaceuticals or primary-care products by interviewing physicians to find out whether specialists or PCPs were the primary prescribers. Taking the most successful quartile (36 drugs, five-year NPV of US sales  $\geq$  \$900 million, median fifth-year US sales  $\sim$  \$590 million) and the least successful (36 drugs, five-year NPV of US sales  $<$  \$140 million, median fifth-year US sales  $\sim$  \$20 million), we assigned scores to each drug according to various measures (see sidebar, “The drug rating system”).

- Differentiation versus standard of care at time of launch in terms of efficacy, safety, and convenience
- Market size at launch
- Five-year market growth rate
- Number of competing drugs at launch
- Degree of novelty in the mechanism of action (MoA)
- Disease severity and unmet need.

We assigned these scores on the basis of interviews conducted with about 50 expert physicians, consultation of the primary medical literature, and information

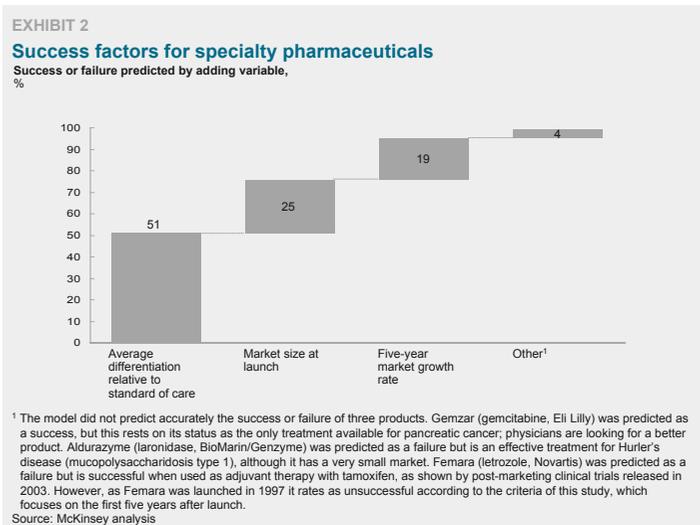
drawn from various commercial databases (Pharmaprojects, EvaluatePharma, and IMS). We then decided which of these parameters correlated with commercial success using the technique of logistic regression.

## Differentiation is the key

Our analysis shows that only three of the above factors were required to explain the commercial success of 96 percent (69 out of 72) of the specialty drugs we examined (Exhibit 2).

The most important is differentiation relative to the standard of care in terms of efficacy, safety, and convenience. This accounts for 51 percent of a specialty drug's success. In particular, the top 36 products were all significant improvements on the treatments available at the time of launch. Successful products did not have to improve on all aspects of efficacy, safety, and convenience; many were superior in one respect only and close to parity for the others.

Market size at launch is the second most important factor, accounting for 25 percent of success. Simply stated, it is easier



## The drug rating system

To find out which factors were important in deciding the success and failure of the drugs in our study, we rated each product according to various criteria and then used regression analysis. To give equal weight to all factors for the regression analysis, we rated each product using a five-point scale.

**Differentiation versus standard of care (in terms of efficacy, safety, and convenience) at launch.** 1, substantially inferior; 2, slightly inferior; 3, at parity; 4, slightly superior; 5, substantially superior. Source: interviews with physicians

**Market size at launch of primary indication.** 1,  $<$ \$0.25 billion; 2, \$0.25-1.5 billion; 3, \$1.5-3 billion; 4, \$3-5 billion; 5,  $>$ \$5 billion. Source: EvaluatePharma; McKinsey analysis

**Market's annual growth rate over five years post-launch.** 1,  $<$ 0 percent; 2, 0-10 percent; 3, 10-20 percent; 4, 20-30 percent; 5,  $>$ 30 percent. Source: EvaluatePharma

**Competitive intensity (for example, number of competing products available at launch).** 1,  $>$ 6; 2, 5-6; 3, 3-4; 4, 1-2; 5, 0. Source: EvaluatePharma; The Medical Letter; interviews with physicians; McKinsey analysis

**Novel mechanism of action (MoA).** 1, simple reformulation; 2, complex reformulation (for example, pegylation); 3, novel MoA, five to ten years after lead; 4, novel MoA, two to five years after lead; 5, novel MoA, lead candidates. Source: EvaluatePharma; Pharmaprojects; interviews with physicians; The Medical Letter; McKinsey analysis

**Disease severity/unmet need.** 1, no risk of morbidity; 3, some severity; 5, high morbidity. Source: interviews with physicians

to have a successful product in a large, existing market than to create a market for a new product—which helps to explain big pharma's focus on “me-too” products. This finding is supported by the work of Ramesh Hariharan and Navjot Singh, which showed that market size at launch and differentiation were the only factors correlating with the success of drugs made by large companies.

Market growth rate over the first five years after launch is the third vital factor for success with 19 percent correlation, showing that specialty products have the capacity to be “market creators.” Rituxan (retuximab), Genentech's revolutionary treatment for non-Hodgkin's lymphoma, is an example of a drug that was able to expand a market from about \$20 million in total sales to \$1.5 billion in only five years. To understand this phenomenon of market creation better, we analyzed all the drugs

in our survey that had small markets at launch ( $\leq$ \$250 million in annual sales) and found that the successful ones were highly differentiated (average differentiation of 4.0 on a 5-point scale) with respect to the standard of care at launch and were focused on severe diseases with high unmet need (Exhibit 3). By contrast, products that were unable to differentiate themselves sufficiently from the pre-existing standard of care, with an average differentiation slightly lower than the standard of care (average differentiation of 2.9), were failures.

In addition to understanding which factors influence success, it is equally important to understand those that do not. Although specialty products are often considered highly novel, novelty did not correlate with success. This indicates that a novel MoA does not necessarily amount to differentiation from a physician's perspective.

### EXHIBIT 3 Specialty drugs launched into small markets

Drugs launched 1994-2003 with initial market sizes \$250 million<sup>1</sup>

	Drug	Company <sup>2</sup>	Average differentiation	Market size at launch, \$m	Market five-year CAGR, <sup>3</sup> %	Novelty of mechanism of action	Disease severity
Successful drugs	Rituxan (rituximab)	Genentech/Biogen Idec	4.2	23	132	5	5
	Synagis (palivizumab)	MedImmune	4.6	142	46	5	5
	Aricept (donepezil)	Eisai	4.2	236	37	1	4
	Gleevec/Glivec (imatinib)	Novartis	4.8	257	68	5	5
	Camplosar (irinotecan)	Pharmacia	3.4	72	74	5	5
	ReoPro (abiximab)	Eli Lilly	3.5	23	86	5	4
	Gemzar (gemcitabine)	Eli Lilly	3.2	62	64	1	5
<b>Average</b>			<b>4.0</b>	<b>116</b>	<b>72</b>	<b>3.9</b>	<b>4.7</b>
Unsuccessful drugs	Cervidil CR (dinoprostone)	Forest Laboratories	3.3	18	12	1	1
	Aldurazyme (laronidase)	BioMarin/Genzyme	4.2	12	63	5	5
	ProAmatine (midodrine)	Shire	2.5	230	-2	5	5
	Trisenox (arsenic trioxide)	Cephalon	3.2	138	23	5	5
	Infasurf (calfactant)	Forest Laboratories	3.2	154	0	1	5
	Ontak/Onzar (denileukin diftitox)	Eisai	3.3	28	12	5	5
	Refludan (lepirudin)	Schering AG	2.8	20	46	3	4
	Lustra/Lustra-AF (hydroquinone USP 4%)	Medicis Pharmaceuticals	2.7	11	6	1	1
	Kinerase (kinetin)	Valeant	3.1	31	58	5	1
	Vistide (cidofovir)	Gilead	2.4	221	-4	3	5
	Salagen (pilocarpine)	MGI Pharma	2.7	4	39	1	1
	Cetrorelix (cetrorelix)	Serono	2.8	11	23	5	1
	Carticel (autologous cultured chondrocytes)	Genzyme	2.2	254	11	5	4
	Glidel Wafer (polifeprosan 20 with carmustine implant)	Guilford Pharmaceuticals	1.9	45	52	2	5
	Oncaspar (pegaspargase)	Enzon	3.4	3	11	2	5
	Photofrin/Photobarr (porfimer sodium)	Axcan Pharma	3.0	87	57	5	2
	Ethylol (amifostine)	MedImmune	2.4	22	89	5	2
DepoCyt (cytarabine liposome injection)	Enzon Pharmaceuticals	3.5	4	12	1	5	
<b>Average</b>			<b>2.9</b>	<b>93</b>	<b>13</b>	<b>4.1</b>	<b>4.0</b>

<sup>1</sup> Some specialty drugs were able to create markets through significant differentiation and focus on severe disease areas.

<sup>2</sup> Company or companies that owned the product during first five years after launch.

<sup>3</sup> Compound annual growth rate.

Source: EvaluatePharma; Pharmaprojects; The Medical Letter; physician interviews; McKinsey analysis

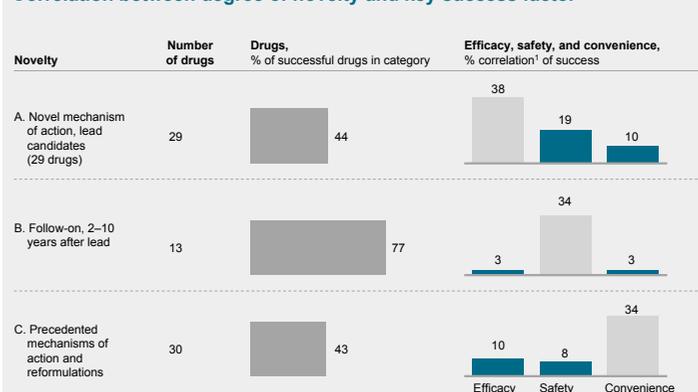
For instance, inspection of the market-creating products (Exhibit 3) shows that successful products did not, on average, have more novel mechanisms than the failures. The number of competing products did not affect success either, illustrating again that it is only a product's relative merits that are significant (that is, its differentiation relative to the standard of care) and not the number of competitors faced.

### First place does not always win

Although our findings indicate that novelty itself is not an indicator of success, we did find that it might provide a way for companies to understand how to position their products.

We segmented all of the products into three groups based on their novelty: first, lead-candidate molecules with novel MoAs; second, follow-on products, launched two to ten years after the lead molecules; and third, precedented mechanisms and reformulations. We then

### EXHIBIT 4 Correlation between degree of novelty and key success factor



<sup>1</sup> Correlation is the R2 of the linear regression of the net present value of the first five years of sales with efficacy, safety, and convenience.

Source: EvaluatePharma; Pharmaprojects; The Medical Letter; physician interviews; McKinsey analysis

looked at the correlation of commercial success with regard to efficacy, safety, and convenience. Exhibit 4 shows that novel molecules are more likely to be successful if they demonstrate improved efficacy, follow-on drugs if they show improved safety, and reformulations if they offer enhanced convenience. The result for follow-on products is particularly striking, as we found not only that an improved safety profile accounts for 34 percent of their success (improvements in efficacy being generally irrelevant, with a 3 percent correlation rate), but also that 77 percent of the products were successful compared with about 40 percent of novel products and reformulations. One implication is that, in certain markets (see discussion below on therapeutic areas), taking time to refine the safety profile and aiming for more desirable indications may be a better strategy than rushing to be first.

### Not all therapeutic areas are equal

Oncology (including immunomodulators), the central nervous system (CNS), and systemic anti-infectives (in particular,

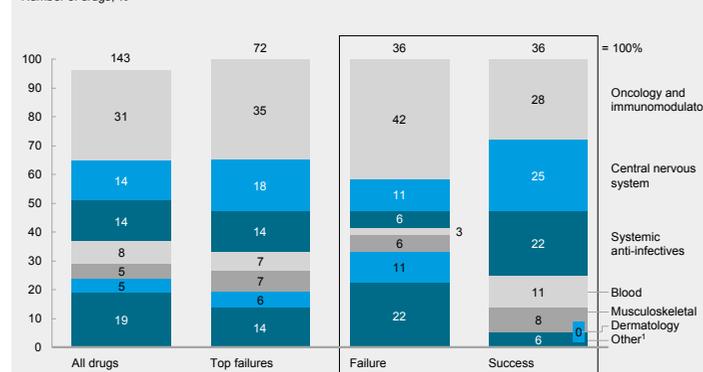
widely used HIV drugs) are the three therapeutic areas that dominate the landscape of specialty pharmaceuticals, representing more than 65 percent of all specialty drugs in our survey (Exhibit 5). However, their respective rates of success and failure are drastically different. Oncology has significantly more (42 percent) of the failures, but only 28 percent of the successes, with success most strongly tied to improvements in efficacy. The large number of oncology failures may be due to the rapidly improving standard of care in oncology, driven by the extreme heterogeneity of cancer and the large numbers of associated pathways and targets, which leads to market fragmentation. For instance, many of the failures ended up as treatments for limited patient populations: Ontak/Onzar (denileukin diftitox, Ligand Pharmaceuticals/Eisai), Trisenox (arsenic trioxide, Cephalon), and Oncaspar (pegaspargase, Enzon) were all indicated for very low-incidence leukaemias or lymphomas.

Also, although oncology has been a hotbed of innovation—Rituxan, Gleevec (imatinib, Novartis) and Herceptin (trastuzumab,

Genentech)—some of the most successful products were based on well-known mechanisms with indication expansions. These include Eloxatin (oxaliplatin, Sanofi Aventis), the first platinum compound to show significant effects against colorectal cancer, and Gemzar (gemcitabine, Eli Lilly), the first pyrimidine analogue with effects on pancreatic cancer, showing that novelty is not a requirement for success in oncology.

### EXHIBIT 5 Dominant therapeutic areas for specialty pharmaceuticals

Breakdown of success and failure by therapeutic area  
Number of drugs, %



<sup>1</sup> Other includes cardiovascular, endocrine, genito-urinary, respiratory, gastrointestinal, and metabolic.

Source: EvaluatePharma; McKinsey analysis

In the field of CNS drugs, specialty products have fared much better, representing 25 percent of successes and only 11 percent of failures. The former are mainly antipsychotics (55 percent), with the rest being treatments for multiple sclerosis, epilepsy, and Alzheimer's disease. Because many CNS-related diseases are chronic conditions, safety played a major role in deciding which products succeeded. In particular, none of the five successful antipsychotics in our survey—Zyprexa (olanzapine, Eli Lilly), Risperdal (risperidone, Janssen), Abilify (aripiprazole, Otsuka/Bristol-Myers Squibb), Seroquel (quetiapine, AstraZeneca), and Geodon (ziprasidone, Pfizer)—was viewed by physicians as having any real efficacy advantage compared with others (all had average efficacy scores of 3.0), but all were perceived as having various safety benefits.

Another interesting CNS success story is Aricept (donepezil, Eisai/Pfizer), which has become the standard of care for Alzheimer's disease. Its efficacy was viewed by many physicians as “unremarkable,” but, owing to a good safety profile and a lack of other treatment options, they also commented that they did not see “any harm” in prescribing it.

In some regards, it is surprising to note the relative success of CNS drugs, a therapeutic area that many companies target less aggressively, owing to the more limited understanding of the underlying biology and hence the narrower set of known “druggable” pathways and targets. For example, the antipsychotics all target some combination of serotonin and dopamine receptors, whereas the antiepileptics mainly target various pathways involving the GABA ( $\gamma$ -aminobutyric acid) receptor. Nevertheless, even given the challenges

associated with identifying new CNS targets, our research highlights that the CNS is a therapeutic area in which the rewards and success rates may warrant more significant exposure.

Anti-infectives/HIV therapies have a similar profile to CNS drugs in that they represent more successes (22 percent) than failures (6 percent). As HIV has been transformed into a chronic disease, so successful therapies have come to be distinguished from earlier drugs by an enhanced safety profile. Reyataz (atazanavir, Bristol-Myers Squibb) and Viread (tenofovir disoproxil fumarate, Gilead) are successful largely because they have fewer side effects than some of the older, comparable therapies. In addition to the safety profile, several HIV drugs—Combivir (lamivudine + zidovudine, GlaxoSmithKline), Trizivir (abacavir + zidovudine + lamivudine, GlaxoSmithKline), and Kaletra (lopinavir + ritonavir, Abbott) were successful owing to the convenience benefits of combination therapies, which reduced the number of pills taken by the patient. Emtriva (emtricitabine, Gilead) provides an interesting example of the benefits of the combination pill to the patient; formulated in isolation it is one of the “failures” on the list because of its “inconvenience.” However, it is also a component, along with Viread, of the highly successful combination pill Truvada (which launched in 2004, so is not included in this analysis).

### Winning strategies

Our research suggests that what really matters to specialty physicians is creating a therapy that is a significant improvement over the standard of care. Novel MoAs might represent exciting science but are not correlated with differentiation from the standard of care, whereas less “exciting” approaches such as reformulating drug

delivery can produce results. We suggest that several possibilities exist to increase the likelihood of success in the specialty market. First, include more practising clinicians in the process of defining the target product profile to bring about a more practical assessment of the true value of a new drug over the likely standard of care at time of launch. Second, encourage early-stage development teams to articulate potential approaches to differentiation during portfolio reviews. Third, ensure the right trade-offs are made (especially between safety and efficacy) in designing clinical trials, particularly for drugs that are not first to market.

An additional question is whether or not these findings will continue to ring true considering the many potential changes in the pharmaceutical landscape in the next few years. With the rise in personalized medicine, increased pressure from payors for “pay-for-performance” therapies, and the probable greater use of pharmacoeconomic analysis, the hurdles for new drugs are likely to multiply rather than diminish. If this proves to be the case, then the need to create differentiated products—and knowing how to surpass the future standard of care—will be more important than ever.

A version of this article, entitled “What drives success for specialty pharmaceuticals?” was first printed in *Nature Reviews Drug Discovery*, Volume 7, pp. 563–567 (June 6, 2008). For further details of the methodology and analysis, please contact the authors.

**Edd Fleming** ([edd\\_fleming@mckinsey.com](mailto:edd_fleming@mckinsey.com)) and **Philip Ma** ([philip\\_ma@mckinsey.com](mailto:philip_ma@mckinsey.com)) are directors in McKinsey's Silicon Valley office, where **Laura Furstenthal** ([laura\\_furstenthal@mckinsey.com](mailto:laura_furstenthal@mckinsey.com)) is a principal. **Mark Gudiksen** is an alumnus of the San Francisco office.

The authors wish to thank Jay Chiang, Jerel Davis, Padma Sundar, and Denise Flaherty for their assistance in preparing this article.

## R&D in emerging markets: a new approach for a new era

Gone is the time when R&D in developed markets could meet the pharmaceutical needs of emerging markets. Local presence is required.

Michael Edwards

The pharmaceutical industry has long been driven by the demands of the north American and west European markets, with scant attention given to emerging markets and the diseases more prevalent there.

This is changing. Emerging markets contributed 30 percent of the pharmaceutical industry's value in 2008<sup>1</sup> and are forecast to grow by 14 percent a year to 2013.<sup>2,3</sup> By contrast, the US market—which still represents 40 percent of the global market—is expected to grow more slowly with annual growth of less than 3 percent over the next four years. Indeed, all developed markets tend to be characterized by low or negative growth, a stricter regulatory environment, increasing patent litigation, and greater scrutiny of drug safety.

Not surprisingly, many pharmaceutical companies now see some of the largest emerging markets as the key to their growth ambitions. Over the past ten years, they have established low-cost research institutions in emerging markets, tapped into talent pools there in search of innovation, and set up corporate social responsibility programs to help tackle neglected diseases. But more needs to be done if companies are to capture the opportunity.

Most still tend to regard emerging markets mainly as a source of lower-cost R&D.

Instead, they should consider how R&D can best serve these disparate and often poorly understood regions. The likely outcome is that they will place a significant proportion of their scientists, their chemistry, manufacturing, and controls, and their development groups there. Only in this way will they understand the needs and preferences of patients and physicians, develop products that meet those needs, and ensure the products are introduced in a timely manner.

### Medical needs and preferences

Patients in emerging markets have different medical needs from patients in the West, and they and their physicians have different treatment preferences. The result is demand for products that reflect these variations. This affects decision making and resource allocation across pharmaceutical companies' entire R&D value chain, from basic research to in-market investment. There are several areas to consider:

### Choice of disease area

The prevalence of some diseases varies significantly between geographic regions and ethnic groups. This can be due to genetic differences or environmental factors such as diet and living conditions. Diseases that mainly affect developing countries and that are not prevalent in



developed markets have been largely ignored from a pharmaceutical “for profit” viewpoint. However, as economies in emerging markets grow and governments begin paying for treatment for the poor, certain hitherto neglected types of disease present revenue opportunities.

The better to understand these diseases, companies are realizing that they need dedicated R&D resources in emerging markets, as valuable insights from patients, physicians, and payors can be gained only through local presence. Currently, oncology is the therapeutic area with the most significant commercial opportunities for drugs developed specifically for emerging markets. Hepatocellular carcinoma (HCC), often caused by hepatitis B and C virus (HBV/HCV) infection, is the most frequently cited example. HCC is the fourth most common cancer globally, and 75 percent of the one million people affected each year live in east Asia. Other cancers more prevalent in developing countries include Kaposi’s sarcoma—less frequently seen in developed markets since the advent and use of the HIV “triple cocktail”—and gastric/oesophageal cancers, which have much higher prevalence in east Asia. HBV infection represents another opportunity: of the 400 million people infected with HBV worldwide, roughly one-third are in China. Many pharmaceutical companies already focus on some of these opportunities. Novartis, for example, is investing in an R&D center in Shanghai, which will initially focus its research on diseases particularly common in China (including hepatitis B and C).<sup>4</sup>

Other neglected diseases include TB, malaria, roundworm, and Chagas. Although these diseases are a major health problem across large parts of the world, they are not among the 20 most prevalent diseases in the largest emerging

markets, with the exception of TB in India.<sup>5</sup> Their value, while growing, will remain small in the near term by developed market standards, though some resources are being applied to dealing with TB. AstraZeneca, for example, has established an R&D center in Bangalore dedicated to TB. Research into these diseases is often conducted in conjunction with public and charitable bodies, or in the context of access-to-medicine programs.

In the case of “diseases of poverty” such as the dehydrating illness cholera, it is not so much research that is needed as better basic living standards. For sufferers of cholera, the treatment is as simple as having access to clean drinking water and rehydration powder.

#### Efficacy and dosing

Dosing levels may need to be adjusted in emerging markets to reflect variations in the efficacy and toxicity of a drug at a given dose between different geographical or ethnic populations. For example, some subpopulations or ethnic groups have lower levels of critical P450 enzymes, which affect drug metabolism and therefore dosing requirements.

Body mass also affects dosing levels, and populations in emerging markets tend to have lower average body mass than those in the United States or Europe. In Japan, local regulators increasingly approve lower doses than their counterparts in the United States. Approved doses for Bayer’s Ciprofloxacin, for instance, vary by region—600 mg in Japan, up to 800 mg in the United Kingdom, and up to 1,200 mg in the United States.

Genetic differences between certain ethnic groups are a further factor in dosing levels—although personalized medicine,

whereby the dose is tailored to genetic differences between individuals rather than groups, is still some way off. Thirty percent of Asians, for example, have a variation in their cytochrome P4502C19 gene which results in reduced ability to metabolize up to 15 percent of all clinically useful drugs,<sup>6</sup> compared with 6 percent of Caucasians. This influences clinical practice. Physicians in Hong Kong commonly prescribe lower doses of certain drugs—diazepam (Valium) is one—for Chinese patients than for Caucasians, because individuals carrying this variant are at higher risk of experiencing toxicities when taking these drugs. In addition, the efficacy of some drugs is limited to certain genetically defined populations or ethnicities, and as a result regulators might approve a drug only for particular ethnic groups. BiDil—a fixed-dose combination of isosorbide dinitrate and hydralazine used to treat heart failure—was dismissed when it failed to demonstrate efficacy in a heterogeneous population trial, but revived when efficacy was shown in patients with significant African ancestry.<sup>7</sup>

As with disease areas, real insight into the nuances of dosing can be gained only through close interaction with local experts and local experience.

#### Patient and prescriber preferences

Understanding physicians’ and patients’ preferences and responding to them effectively is important to success in emerging markets. Fixed-dose combination medicines—drug therapies with two or more active pharmaceutical ingredients (APIs) combined into one tablet—are preferred in many markets, but their popularity still varies greatly by individual market. Their benefits are that they can be more convenient for the prescribing

physician and the patient, and they can improve patients’ compliance.

The polypill, manufactured by India’s Cadila Pharma, includes five APIs and is one of the most extreme examples of a fixed-dose combination. The polypill, which is in Phase III, will be targeted at patients with elevated blood pressure or high cholesterol, and combines hydrochlorothiazide, atenolol, ramipril, and simvastatin, with low-dose aspirin and folic acid.<sup>8</sup> The polypill therefore includes APIs originally made by AstraZeneca, Sanofi-Aventis, and Merck (although all the drugs are now available as generics). Already on the market is the CV Pill, made by Torrent, which combines atorvastatin, ramipril, metoprolol, and aspirin. In some markets these combinations can be priced at a premium and win patent protection.

Formulations represent another opportunity to meet local needs while also potentially extending a product’s lifecycle. There are many formulations that can be considered, including extended-release, soluble powders, inhalers, nasal sprays, injectables, and long-acting patches. In each case, the benefits of meeting an emerging market need or preference have to be weighed against the cost of developing a novel formulation, and thought given to whether the formulation can be developed in time to extend the lifecycle adequately.

The logistics of delivering products to hospitals and pharmacies is another important area that pharmaceutical companies will need to understand. For example, heat and humidity can destroy medications, and there may not be a reliable cold-chain supply system to deliver medication from factory to hospital. In 2007, in Uganda, Abbott launched

Aluvia, a heat-stable tablet formulation of the protease inhibitors lopinavir and ritonavir. The soft-gel capsules and oral solution that were previously available were not appropriate due to the scarcity of refrigeration equipment. The tablet version has now been filed for registration in more than 150 countries, and nearly 75,000 patients were treated in 2008.

## Coming to market

### Regulatory requirements and the move towards global trials

Establishing R&D activities in emerging countries will help pharmaceutical companies obtain market access in a timely way. To a certain extent, some local presence is already a prerequisite. China, India, South Korea, and Taiwan all require trials with local patients for local product registration. Some pharmaceutical companies use contract research organizations to meet these requirements, but others have established their own development centers to run the trials. Sanofi-Aventis, for example, established a Biometrics Research Center in Beijing in 2008 to support global and local clinical trials, from the initial design of studies through to data management and statistical analysis.

Local requirements apart, there are other benefits of including emerging markets in trials for new drugs. Development cycles can be reduced owing to faster recruitment of subjects from a larger pool of patients. And the costs of recruiting patients and paying investigators are lower. This can benefit both the development of drugs that are specifically targeted at emerging markets and that of drugs for developed markets, as patients recruited for trials in emerging countries can count towards

the total patient numbers required to satisfy US and European regulators.

By including relevant populations from emerging markets in trials, companies might also be able to reduce the current time delay that exists between launching their drugs in the West and in emerging markets—a delay that can mean the loss of significant patent revenue. Some companies are beginning to tackle this issue. Merck, for example, has made a public commitment to launch its drugs simultaneously in both spheres—no small undertaking given that some of its previous launches in emerging markets have been up to ten years behind those in the West. Other companies also appear to be moving towards this goal: Bayer was selling the multi-targeted kinase inhibitor sorafenib (Nexavar) for HCC in China less than a year after its US launch, while Bristol-Myers Squibb launched the antiviral HBV therapy entecavir (Baraclude) in China with only a six-month delay.

### Market shaping through local R&D

Conducting R&D in emerging economies provides a number of market-access benefits to pharmaceutical companies by engaging local stakeholders early in drug development.

Working with leading physicians at this stage can help build a product's name and, most importantly, give companies insights into how to tailor products to local needs. For example, working with local physicians in east Asia to develop new oncologic drugs (such as tyrosine kinase inhibitors) could help pharmaceutical companies learn how local oncologists approach cancers such as GI stromal tumours, which are rare in developed countries.

A local R&D presence also helps build a company's reputation for innovation

and attentiveness to specific needs. In countries that do not require inclusion of local patients as a condition of regulatory approval, local Phase IV post-launch trials or epidemiological studies can help demonstrate efficacy in the immediate patient population. (The number of Phase IV trials in emerging markets has grown at an annual rate of more than 50 percent in the past five years.) Post-launch activities, such as Phase IV trials in new indications, might also lead to expanded indications, even on a local basis, providing the sales force and medical scientific liaison staff with additional scientific data for their discussions with local doctors, and reinforcing the message that the company is focused on meeting the needs of the regional market.

Local R&D activities also help build relationships with governments and regulatory agencies. In some places, a local presence is already a prerequisite for market access, as is increasingly the case in Russia; elsewhere, it helps to establish relationships with the authorities that might accelerate approval of a drug and deliver higher reimbursement.

By building significant R&D resources in emerging markets, pharmaceutical companies will revolutionize the global R&D group, helping shift the focus of the entire development pathway—from early development to lifecycle management—to include both developed and developing markets.

For the pharmaceutical industry it will mark the end of an era when it was assumed that the needs of the world's pharmaceutical market were largely those of developed countries. The new era will require considerable organizational change. R&D departments will need new resources. And since emerging market revenues are likely to be less than half those of the United States and European Union in the medium term, funding might have to be ring-fenced. But perhaps the most important change of all, given the speed at which the market is developing, will be a rapid change in mindset. Companies already recognize the need to tailor their product portfolios for emerging markets, but to be real leaders in the future they must tailor their R&D activities too.

Michael Edwards (*michael\_edwards@mckinsey.com*) is an associate principal in McKinsey's London office.

<sup>1</sup> IMS World Review, 2009.

<sup>2</sup> IMS Health, Market Prognosis, March 2009.

<sup>3</sup> "Global pharmaceuticals: emerging markets—infusing badly needed revenues for years to come," Bernstein Research, May 2009.

<sup>4</sup> See <http://www.novartis.com/research/pharmaceutical.shtml>.

<sup>5</sup> "Mortality and burden of disease estimates for WHO member states in 2004," World Health Organisation, February 2009.

<sup>6</sup> "Racial differences in response to pharmacological treatment," Chung, A, 2004.

<sup>7</sup> Abdallah S. Daar and Peter A. Singer, "Pharmacogenetics and geographical ancestry: implications for drug development and global health," *Nature Reviews*, 2005, Vol. 6, 2005.

<sup>8</sup> "Effects of a polypill (Polycap) on risk factors in middle-aged individuals without cardiovascular disease (TIPS): a Phase II, double-blind, randomized trial," The Indian Polycap Study, *The Lancet*, 2009, Vol 373.

## The microeconomics of personalized medicine

Personalized medicine promises to increase the quality of clinical care and, in some cases, decrease health care costs. The biggest hurdles are economic, not scientific.

Jerel Davis, Laura Furstenthal, Amar Desai, Troy Norris, Saumya Sutaria, Edd Fleming, and Philip Ma

Personalized medicine—defined here as a tailored approach to patient treatment based on the molecular analysis of genes, proteins, and metabolites—has generated much excitement. Yet few personalized medicine tests have achieved high clinical adoption. To understand better the challenges to the development and acceptance of personalized medicine, and how to overcome them, we interviewed more than 60 leading payors, providers, regulatory experts, pharmaceutical and biotechnology companies, academic opinion leaders, and diagnostics and clinical laboratory companies, and conducted microeconomic analyses of different stakeholder issues (see sidebar, “Research details”).

Our investigation highlighted three main obstacles to the advancement of personalized medicine: first, scientific challenges (poor understanding of molecular mechanisms or a lack of molecular markers associated with some diseases, for example); second, economic challenges (that is, poorly aligned incentives); and third, operational issues (electronic tracking of diagnostic information, privacy concerns, reimbursement coding problems, provider/patient education). Although scientific difficulties remain, it now seems that the economic challenges and operational questions are the biggest hurdle. In many cases, operational issues can be largely resolved within a particular stakeholder

group. However, correcting the incentive structure and modifying the relationships between stakeholders could be more complex.

In this article, we discuss the economic challenges of personalized medicine from the perspective of four key groups of stakeholders: payors, providers, pharmaceutical and biotechnology companies, and diagnostics companies. We focus on the US market

EXHIBIT 1  
Not all diagnostic tests save costs for payors<sup>1</sup>

	Companion diagnostics			Procedure-focused diagnostics	Genetic risk markers		
	Her2 <sup>2</sup>	BCR-ABL <sup>3</sup>	Warfarin	AlloMap	BRCA1	BRCA1-F <sup>6</sup>	KIF8 (statin)
Savings from changed decision, \$ thousands	40	80	2	4	25	25	-3.5
Probability that diagnostic changes treatment decision, %	70	5	35	75	2	20	50
Savings per test, \$ thousands	28	4	0.7	3.1	0.5	5	<0
Cost of test, \$ thousands	0.1	1	0.3	3	2-3	2-3	na
Cost saving for payors?	✓	✓	✓	✓	✗	✓	✗

<sup>1</sup> Estimated savings per test is the product of savings from a single changed treatment decision and the probability that any given patient will have a “positive” test (such that treatment decision is changed).

<sup>2</sup> Human epidermal growth factor receptor 2.

<sup>3</sup> Breakpoint cluster region—abelson tyrosine kinase.

in particular, but the matters raised are also relevant elsewhere. Our proposals for overcoming these challenges could significantly accelerate the adoption of personalized medicine.

## Stakeholder incentives and challenges

### Payors

Investors and analysts have suggested that personalized medicine can dramatically reduce health care costs and help payors market products to the most attractive customers. Yet most payors have been slow to invest in personalized medicine. Leaders in payor organizations identify several factors that could explain this reluctance. The first is an inability to identify easily which tests truly save costs. The second is apprehension that it is difficult to track much of the earlier stage/experimental testing, leading to fears that although individual tests may not be that expensive, the overall eventual costs could be unjustifiably high. A third concern is difficulty in enforcing standard protocols to ensure that physicians follow through with appropriate patient care based on test results. Fourth, there is potential for misuse of test information, particularly in the early stages of test investigation/development, which could lead to patient harm. Fifth, there is a lack of longitudinal accounting that would enable payors to capture long-term cost savings from near-term testing.

To understand which tests actually save costs, we analyzed various types of test (see sidebar, “Additional notes,” note 1).

Two primary factors determine a test’s cost effectiveness from a payor’s perspective: per patient savings (that is, the difference between the cost of treating the disease and the cost of the intervention indicated by the test); and the likelihood that a test suggests an intervention for any particular patient (Exhibit 1). Tests that help avoid the use of expensive therapies (for example, trastuzumab [Herceptin, Genentech/Roche] or imatinib [Gleevec, Novartis]), minimize costly adverse events (such as the warfarin dosing test), or delay expensive procedures can be extremely cost effective for payors. Although such tests cost between \$100 and \$3,000 each, they save \$600 to \$28,000 per patient. By contrast, tests that save a small amount per patient or have a low probability of identifying patients requiring intervention are not cost effective. For example, although BRCA1 testing to predict the risk of breast cancer can save around \$25,000 per patient identified, mutations are so rare in the general population that this test, which costs up to \$3,000 per patient, is cost effective only when performed on a patient with a family history of breast cancer. Some tests could also create costs on a per patient

basis. Variants in KIF6, for example, have been linked to a 50 percent increase in the risk of myocardial infarction, but this risk can be reduced to normal levels through treatment with statins.<sup>1,2,3</sup> Widespread use of a hypothetical test based on these markers could actually result in higher costs through treating patients with statins, compared with the savings from avoiding cases of myocardial infarction.

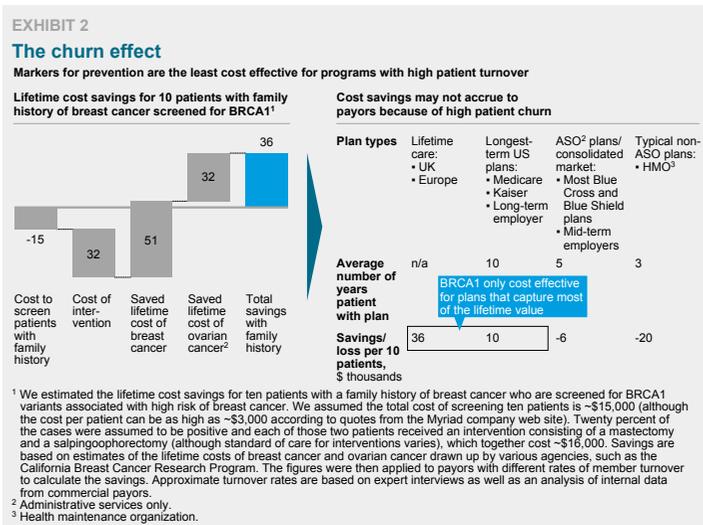
Payors’ adoption of personalized medicine tests is further complicated by the high customer turnover experienced by many commercial payors in the United States. This high turnover makes it less economically attractive for payors to reimburse prophylactic tests that minimize the likelihood of conditions that will occur much later in life. Costs accrue to the payor that screens the patient and performs the intervention, but the benefit accrues to the payor covering the patient when the disease actually arises (perhaps ten years later). The pharmacoeconomics for the BRCA1 test illustrate the point (Exhibit 2). This longitudinal accounting issue is particularly acute for diseases with a late or delayed onset, whereby the insurer for the elderly—for example, Centers for Medicare and Medicaid Services (CMS) in the United States—accrues the benefit of interventions that were paid for years earlier by commercial payors. Notably, payor systems that have low patient turnover, such as integrated systems like Kaiser Permanente in the United States or single-payor systems in Europe, are less exposed to this incentive challenge.

As described above, personalized medicine tests span a spectrum from cost effective to cost creating. Because the actual cost savings may not be known until the test has been on the market for some time, it will remain in payors’

interests to delay adopting personalized medicine tests until they can differentiate between those that are cost saving and others that are cost creating. The winning strategy for diagnostics companies may therefore be to collaborate with other stakeholders where the economics are more aligned (for example, Kaiser Permanente, large self-insured employers, and the Veterans Affairs system in the United States, all of which have lower membership turnover). Generating high-quality health economic evidence will provide the reimbursement confidence that enables payors more rapidly to adopt tests and align physicians’ incentives with patient care and outcomes, rather than procedures. This could create a source of competitive advantage for payors that are more successful in identifying and implementing policies to promote cost-saving diagnostics.

### Providers

Today’s “procedure-based” reimbursement system for providers also presents a challenge to the adoption of personalized medicine. In this system, provider economics will create incentives for the use of some personalized medicine tests, but might discourage the use of others. Physicians could be more likely to embrace tests that increase the number of procedures performed, while taking a more hesitant approach to those that diminish procedure volume. For example, a test that identifies patients at high risk of colon cancer such that they require colonoscopies at three times the normal frequency would align well with gastroenterologists, given that the lifetime value of a patient related to such a molecular diagnostic is around \$2,000. Other tests may be cost neutral, or have negative microeconomic incentives for their use. For example, Oncotype Dx, a gene-based diagnostic test for breast



cancer patients that can be used to assess the likelihood of benefit from chemotherapy, ultimately reduces the number of patients that physicians treat with such chemotherapy, and thus the revenue that those patients generate. Even so, Oncotype Dx has been widely adopted because of its clinical merit, but this example illustrates the challenges that such tests can pose to providers' economics.

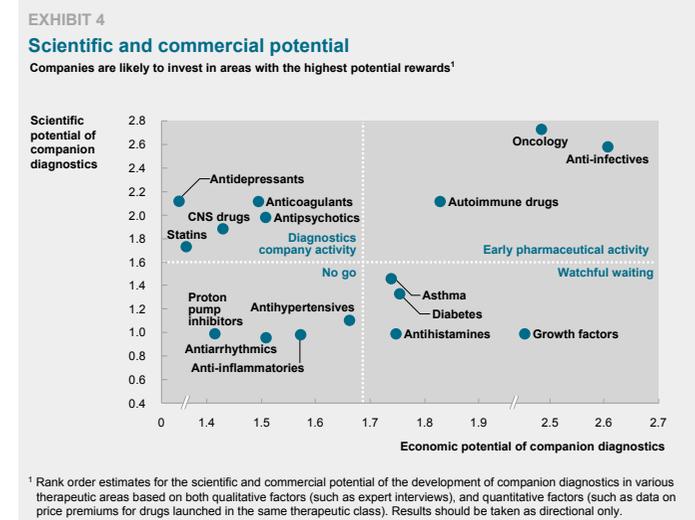
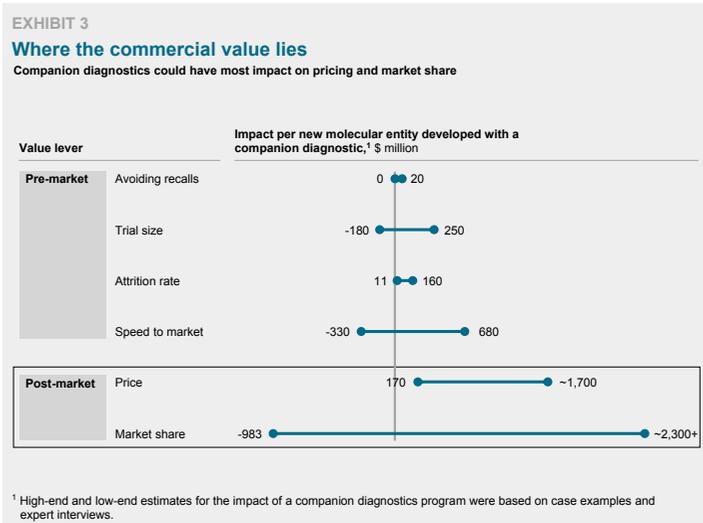
### Pharmaceutical and biotechnology companies

These companies are now using biomarkers to aid R&D, and in some cases will develop these markers as companion diagnostics (tests to identify patients' likelihood of responding to a drug or experiencing side effects). R&D executives at 16 of the top 20 biopharmaceutical companies interviewed in a survey by McKinsey in 2007 indicated that, on average, 30 to 50 percent of drugs in development have an associated biomarker program, and suggested this number was likely to increase. By contrast, the same executives also suggested that fewer than 10 percent of drugs with current biomarker programs would be launched with a companion diagnostic over the next five to ten years (and this is highly dependent on the disease area).

In theory, companion diagnostics can improve R&D productivity by decreasing trial size, reducing attrition, and/or increasing speed to market, and enhance commercial performance by boosting market share and/or supporting higher drug prices. However, many

companies are moving slowly towards the application of biomarkers and companion diagnostics. This is evidenced by the fact that while the most aggressive players have biomarker programs for 100 percent and companion diagnostics for 30 percent or more of their compounds, the average company has far fewer (30 to 50 percent and less than 10 percent respectively). Moreover, many of the experts we interviewed stated that their corporations had not prioritized companion diagnostics and were taking a "cautious approach" to investments. Scientific and clinical factors pose some limitations to the pace of development. In some disease areas, understanding of molecular mechanisms is insufficient to select biomarkers rationally at early stages of development. In other areas, there is not a large clinical need for companion diagnostics. However, in many disease areas, companies are moving slowly despite scientific advances.

Our research suggests that the potential to generate greater value after marketing through increasing price and market share is vastly more important for the economics of pharmaceutical and biotechnology companies than improving development



productivity. Indeed, it seems companion diagnostics may do little to improve development productivity. In many cases, they might actually increase overall cost and delay development. With respect to clinical trials, experts suggested that Phase II trials often have to be larger when companion diagnostics are employed. In practice, trials often need to be designed with several potential candidate biomarkers in Phase II (and sometimes Phase III) as it is unclear which markers will be predictive (see "Additional notes," note 2). In addition, the US Food and Drug Administration (FDA) is likely to require that "marker-negative" patients be included in Phase III trials, based on concerns that the drug could be used off-label in these patients. This practice is likely to eliminate the upside from smaller trials that has been widely cited in recent years (see "Additional notes," note 3). Apart from trial size, other commonly cited applications of personalized medicine during drug development also seem unlikely to improve drug development productivity substantially (Exhibit 3).

Although increasing development productivity may not provide sufficient

incentives for companies to pursue companion diagnostics, there are significant potential commercial benefits from increased market share and pricing power. At the same time, there is also significant risk, as companion diagnostics divide the treatable patient population into sub-segments and can reduce market share in some cases. Given this, companion diagnostics are most likely to be value creating for later-to-market

entrants in crowded markets characterized by significant pricing flexibility.

For example, if two drugs are already on the market and are relatively undifferentiated, the third drug on the market is likely to capture a relatively small share—say, 5 to 20 percent (see "Additional notes," note 4). A companion diagnostic that identifies a segment of the patient population that will respond especially well to a drug or have lower toxicity, and thereby enables higher pricing, could generate value. A key determinant of pricing diversity is payors' price scrutiny/sensitivity, which varies dramatically by disease area, particularly in the United States. This is illustrated by Bidil, a fixed-dose combination of two generic cardiovascular drugs, hydralazine and isosorbite dinitrate, that has been approved by the FDA specifically for African Americans with heart failure. In this case, attempts to charge a price premium were met with aggressive differential co-pay tiering by payors, which contributed to lower sales than expected (see "Additional notes," note 5). In therapeutic classes where payors

scrutinize prices less intensely (oncology drugs, for example), companies would be more likely to charge a premium and maintain coverage.

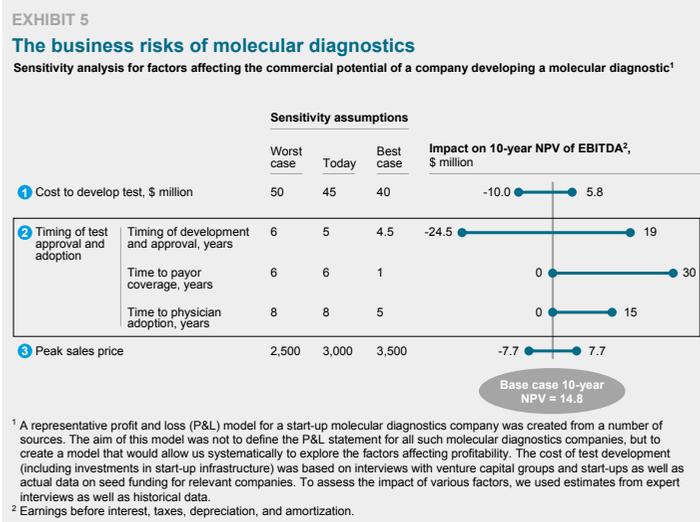
These companies should be, and are, considering investing in personalized medicine in certain disease areas. To highlight disease areas where near-term investment in companion diagnostics is most likely to occur, we segmented drug classes according to their scientific and commercial potential (Exhibit 4). This segmentation is based on quantitative factors, as well as qualitative factors from interviews, and the results should be taken in this light. Our analysis indicates that companies are most likely to invest in diagnostics in areas such as oncology, immunology, and infectious disease. The segmentation also reveals disease areas where technical feasibility and clinical need exist, but where incentives are not aligned to drive investment. These areas, such as anticoagulants, antipsychotics, and antidepressants, are ripe for development by other organizations, such as diagnostics companies.

However, companies should also realize that the payor environment is evolving rapidly and application of personalized medicine tools will increasingly be required to preserve value. Although pharmaceutical and biotechnology companies need to be aware of areas where diagnostics can destroy value by sub-segmenting their existing markets, it will be equally important to prepare for an environment in which regulatory bodies will demand greater proof of patient outcomes

to justify approval, reimbursement, and price. With this in mind, companies should act quickly to build the capabilities and experience that will be needed to succeed.

### Diagnostics companies

Diagnostics and life science tools companies enable a wide variety of test types, including companion diagnostics (often in collaboration with a biotechnology or pharmaceutical company), early-stage diagnostics, disease recurrence and monitoring tests, adverse drug events tests, and genotypic risk marker analyses. However, diagnostic test developers have faced difficulty capturing the full value they generate, as exemplified by the fact that diagnostic tests are estimated to influence between 60 and 70 percent of all treatment decisions, yet account for only 5 percent of hospital costs and 2 percent of Medicare expenditure.<sup>4</sup> Molecular diagnostics are often cited as a more attractive market segment than typical diagnostics, given the potential for higher prices (\$100 to \$3,000 per test compared with \$20 to \$50 for a typical diagnostic test) and higher gross margins (50 to 70 percent for a sample



molecular diagnostic compared with 30 to 50 percent for most typical large laboratory companies [see “Additional notes,” note 6]). Indeed, a number of emerging companies, including Genomic Health, Myriad, Celera, Monogram, and Xdx, have been successful in raising funding and developing innovative molecular diagnostic tests.

Unfortunately, the molecular diagnostics business case still holds significant risk (Exhibit 5). A number of factors contribute to this risk, including development costs, timing of development and approval, time to payor coverage, rate of provider adoption, and peak sales price. To understand the relative importance of these factors, we modeled the economics of a hypothetical start-up esoteric diagnostics company, then performed a sensitivity analysis using upside and downside scenarios for each variable. It should be noted that this model was based on benchmarks from a few molecular diagnostics businesses with the intent of testing the importance of risk factors. The model does not represent a specific company, and the economics for companies with products currently on the market vary significantly. Based on this model, the expected ten-year net present value (NPV) for an average diagnostic test is around \$15 million. The most important factors influencing profitability are the time to approval and rate of payor adoption. If the time to approval is delayed by a year, the ten-year NPV becomes negative at around -\$10 million. This finding is relevant given that it remains unclear how the FDA will regulate *in vitro* diagnostic multivariate index assays (IVD-MIA). At the time of writing, the FDA suggested a 510(k) may be sufficient for tests which are prognostic indicators (see “Additional notes,” note 7), but a premarket approval (PMA) from the FDA is likely to be required if the test directly influences therapy decisions. PMA review

is likely to increase the time to market by at least a year (see “Additional notes,” note 8). That said, good communication between the Center for Drug Evaluation and Research and the Office of In Vitro Diagnostic and Device Evaluation and Safety may partially mitigate this for priority reviews. It remains unclear what the approval timelines for other systems will be. The European Medicines Agency (EMA) and Japan’s Pharmaceuticals and Medical Devices Agency (PDMA) also have yet to establish clear guidelines for how they will treat the approval of personalized medicine tests.

With respect to payors’ adoption, the case of Oncotype Dx demonstrates the challenge of slow coverage. Although the test was launched in 2004, analysts and company estimates suggest it will be 2010 before all payors routinely cover the test. Coverage stands at about 85 percent. This contrasts starkly with typical adoption of a new drug. In the United States, new drugs are generally reimbursed immediately at launch or within the year. In Europe, drug coverage may take slightly longer, depending on the extent of the review, but is unlikely to take more than four years—the adoption timeline experienced by Oncotype Dx.

Start-up diagnostics companies therefore face challenging economics. However, as more tests become available and payors, regulators, and molecular diagnostics companies gain experience, development and adoption times are likely to shorten. Likewise, as the regulatory process becomes clearer—but potentially longer—payors’ adoption rates may also increase. Given payors’ trepidation about personalized medicine testing, it will therefore be advantageous for leading diagnostics companies to help shape the development of rigorous but efficient regulatory and approval standards.

## Potential catalysts for personalized medicine

We have described how current market failures limit the speed of adoption of personalized medicine, and how solutions to these economic challenges represent opportunities to accelerate market development. On the basis of conversations and analyses conducted during the course of this investigation, we see four main catalysts that could significantly affect the adoption of personalized medicine in the near term.

1. Maximizing transparency and efficiency of the regulatory-approval process
2. Increasing the pace and predictability of payor coverage for appropriate tests
3. Aligning reimbursement practices to incentivize appropriate diagnostic use by physicians
4. Encouraging pharmaceutical and biotechnology companies to take a long-term investment view.

### Regulatory environment

First, regulatory bodies such as the FDA must improve the clarity and efficiency of the test regulatory approval processes, both for stand-alone and companion diagnostics. These clarifications are critical to diagnostics companies' ability to plan ahead and to design trials. Based on our conversations with more than 60 experts, the key questions that regulatory bodies such as the FDA and EMEA should address include: Will marker-negative patients be required for Phase III trials? Will use of archived samples or "flexible" trial designs be permitted for approval of companion diagnostics, and under what circumstances? What regulatory standards and oversight will be required

for personalized medicine tests, especially laboratory-developed tests, to be used in therapy decisions?

For the new regulations under consideration, authorities need to weigh short-term costs against long-term benefits. Current plans include classifying tests as Class I, II, or III, based upon the level of risk of the intended use. IVD-MIA changes that promote more rigorous evaluation of safety and effectiveness may have long-term benefits by encouraging faster adoption by payors and physicians owing to the higher approval standards. However, the near-term consequences may harm short-term market investment.

For diagnostics companies, the approval process can actually be an opportunity to justify higher value/pricing by showing willing to set appropriately stringent standards, and by shaping regulatory guidelines to bolster the industry and protect patients. For its part, the FDA should work to minimize approval delays that result from higher standards, and help mitigate any negative impact on investment in development. Leading pharmaceutical, biotechnology, and diagnostics companies should look for opportunities to help shape the development of these guidelines and standards.

To drive changes in market incentives, regulatory bodies such as the FDA and EMEA could decide not to require collection of clinical data on marker-negative patients, thus lowering development costs. Concerns about the use of therapeutics in the "marker-negative" population could be reduced by parallel moves by payor organizations and regulatory bodies to increase barriers to "off-label" use. Furthermore, regulatory bodies could increase the flexibility of trial design and even allow for the approval of

## Research details

*To gather stakeholders' perspectives on personalized medicine we conducted 60 interviews in the first half of 2008 with executives and opinion leaders from leading private payor organizations, academic research institutions, health care provider organizations (such as academic medical centers and hospitals), regulatory bodies, biopharmaceutical companies, molecular diagnostics and clinical laboratory companies, and venture capital funds. The interviewees were:*

- *Eight payor executives including individuals from private payors (for example, Blue Cross/Blue Shield and Health Net) and CMS. The expertise of these individuals spanned coverage decisions and health technology assessment.*
- *Twenty biopharmaceutical executives in positions ranging from vice-president to chief executive officer. Expertise among these individuals spanned business strategy and operations, R&D, regulatory affairs, and reimbursement.*
- *Thirteen diagnostics executives from large clinical laboratory companies and small and mid-sized molecular diagnostics companies. All interviewees were senior executives with long experience of the diagnostics industry.*
- *Six researchers from leading academic institutions in the United States and the United Kingdom who are recognized as experts in molecular genetics, pharmacogenomics, bioinformatics, and molecular and protein diagnostics.*
- *Three venture capitalists from leading firms that focus on molecular diagnostics investments.*
- *Two attorneys with legal expertise spanning intellectual property, FDA regulation, and health care law.*
- *Eight regulatory experts from the Department of Health and Human Services, the FDA, and NICE.*

*We asked each interviewee's opinions on the challenges and opportunities in personalized medicine for all stakeholders, and discussed in more detail the use of personalized medicine in their own field of expertise both currently and over the next five years.*

*Details of the quantitative analysis and financial modeling we conducted to understand specific stakeholder issues are highlighted in the exhibits.*

companion diagnostics on the basis of retrospective tests of the Dx marker (that is, performed on archived samples). Finally, governments and regulatory bodies could reward the development of companion diagnostics directly by increasing the patent life for drugs developed with companion diagnostics, providing tax-based incentives, and continuing to award grants for R&D.

### Payor coverage

In the United States, approval and reimbursement coverage decisions represent two discrete processes with minimal coordination between the FDA and CMS. Uncertainty remains about

how this coordination will work in other parts of the world and processes have not been established—for example, at the time of writing, the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom had not reviewed a molecular diagnostics test. State payors, private payors, and diagnostics companies can help fuel growth in the personalized medicine market by coordinated efforts to improve the pace and process for coverage decisions. One step could be for CMS to take a lead in aligning the reimbursement process with the regulatory approval process. Pre-submission meetings to delineate data requirements for regulatory and coverage

approval and ongoing joint reviews can facilitate interagency collaboration. Optimal alignment across the two agencies would imply that if suitably stringent guidelines are set, then CMS would provide coverage and adequately reimburse those who meet those hurdles. For example, the requirement of additional health economic data and/or regulatory approval for clinical claims may be reasonable prerequisites for coverage, and could thus help ensure adequate reimbursement, pricing, and value for diagnostics players.

Development of formal guidelines could improve the transparency and efficiency of decisions relating to coverage. Today, CMS typically makes coverage decisions for molecular diagnostics at the regional rather than national level. As a consequence, decisions are made many times based on different guidelines and processes and often with differing outcomes (see “Additional notes,” note 9). Private payors also lack clear guidelines for these decisions. Both CMS and private payors have an important role to play in shaping coverage and payment decisions. Private payors we interviewed are waiting to understand and potentially follow CMS coverage policies (as often occurs with therapeutics).

One way to improve coverage guidelines in both systems and processes would be to establish an agency to assess the clinical and cost effectiveness of tests. This agency could be a coordinated effort by payors, CMS, interested pharmaceutical and biotechnology companies, and diagnostics players, and could take the form of a third-party, non-profit agency, a consortium, or a new government agency. The formation of new oversight agencies (for example, an FDA center for diagnostics) could also help in this regard. Notably, single payor systems such as those that

predominate in Europe are at an advantage when it comes to the adoption of personalized medicine, in two ways. First, they are not as susceptible to longitudinal accounting issues. Second, coverage decisions can be less complex and involve fewer decision makers.

### Physician incentives

Beyond improvements associated with regulatory approval and formal coverage, aligning physicians’ incentives could further hasten adoption. Reimbursement schemes in many countries remain largely “activity based,” with physicians receiving disproportionately higher rates for procedure-oriented services than for evaluation and management activities. As such, there is little financial incentive for physicians to perform tests that might prevent the need for further treatment. In fact, there may be a real financial disincentive.

Efforts are under way to shift towards a more “outcome-based” approach to reimbursement, a system that will provide incentives for physicians to use and act on appropriate personalized medicine diagnostics. Yet payors should also work to develop a system that ensures physicians are reimbursed for the test itself in order to encourage adoption. Moreover, personalized medicine tests today are billed in the United States by “CPT code stacking,” whereby a multivariate assay is billed by adding multiple, generic codes—for example, for a diagnostic based on a single gene (see “Additional notes,” note 10). This approach is not scalable and can lead to billing practices in which laboratories game the system. Eventually, individual codes will need to be developed for each molecular diagnostic that are commensurate with the cost and value of the test and provide appropriate reimbursement to physicians.

### Investment by pharmaceutical and biotechnology companies

Pharmaceutical and biotechnology companies should take a long-term investment view. Some already do: leaders we interviewed who have invested most heavily in personalized medicine suggested they are renewing their focus on outcomes and clinical value in the process of drug discovery. They realize that the drugs they are developing today will be entering markets with more competitors, more pricing pressure, and a higher bar for differentiated clinical outcomes. Not surprisingly, these same companies are investing heavily in personalized medicine.

An aggressive move towards value- or outcomes-based pricing by CMS or private payors could greatly increase the financial value of personalized medicine and so the incentive to invest in it. One possibility might be to employ innovative risk-sharing models for drug and diagnostic coverage. For instance, payors could follow the examples in Europe of bortezomib

(Velcade) for multiple myeloma and the interferon-beta drugs for multiple sclerosis; here reimbursement is contingent upon patient outcomes. Similarly, payors could create innovative risk-sharing agreements with diagnostics companies. A test could, say, receive conditional, partial reimbursement for a number of years until the clinical effectiveness was definitively demonstrated (at which point the diagnostics company would be paid in full). The payor limits cost exposure by covering part of the costs for a limited time, while diagnostics companies benefit from early coverage decisions.

Over the next few decades, the development of -omics sciences and supporting technologies will enable the creation of an increasing number of personalized medicine tests. However, the use of these tests could be hampered by poorly aligned incentives for stakeholders. All stakeholders should therefore work together to help reshape these incentive structures and so reap the benefits of personalized medicine.

A version of this article, entitled “The microeconomics of personalized medicine: today’s challenge and tomorrow’s promise,” was first published in *Nature Reviews Drug Discovery*, April 2009, Volume 8, No 4, pp. 279–86. For further details of the analysis and methodology, please contact the authors.

**Jerel Davis** ([jerel\\_davis@mckinsey.com](mailto:jerel_davis@mckinsey.com)) is a consultant in McKinsey’s Silicon Valley office, where **Saumya Sutaria** ([saumya\\_sutaria@mckinsey.com](mailto:saumya_sutaria@mckinsey.com)) and **Laura Furstenthal** ([laura\\_furstenthal@mckinsey.com](mailto:laura_furstenthal@mckinsey.com)) are principals, and **Edd Fleming** ([edd\\_fleming@mckinsey.com](mailto:edd_fleming@mckinsey.com)) and **Philip Ma** ([philip\\_ma@mckinsey.com](mailto:philip_ma@mckinsey.com)) directors. **Amar Desai** ([amar\\_desai@mckinsey.com](mailto:amar_desai@mckinsey.com)) is a consultant in the Los Angeles office. **Troy Norris** ([troy\\_norris@mckinsey.com](mailto:troy_norris@mckinsey.com)) is a consultant in the New York office.

**Additional notes**

*Note 1. Some payors noted that newer thyroid cancer diagnostics have led to a dramatic increase in the incidence of thyroid cancer (a 250 percent increase from 1973 to 2002), but no improvements in mortality.<sup>5</sup> One explanation for the findings is that most of the incremental detection was for papillary cancers with a very good prognosis.*

*Note 2. Experts we spoke to indicated that between two and six markers are typically chosen for a drug's companion diagnostic program. These markers are usually chosen before Phase II, developed in parallel with the Phase II clinical trial, and then tested retrospectively on Phase II participants.*

*Note 3. A widely cited example in this respect is the Phase III trial for the anticancer drug trastuzumab (Herceptin, Genentech/Roche), a monoclonal antibody that targets HER2. The trial included only 470 breast cancer patients and only a marker-negative arm. The expected size of the trial without a companion diagnostic based on HER2 expression levels has been estimated at 2,200 patients (based on a presentation by Arthur Levinson, CEO of Genentech, in October 2003).*

*Note 4. This figure is derived from McKinsey analysis and IMS sales data; the estimated range is based on 5 percent average market share for third-to-market drugs at around one to three years post launch, and 20 percent average market share of second-to-market drugs at around one to three years post launch.*

*Note 5. Bidil was priced at a premium. In interviews with four different insurance companies, payors indicated that they differentially tiered the drug because they did not think the clinical benefit justified the cost.*

*Note 6. These figures are based on McKinsey analysis of selected molecular diagnostics and traditional diagnostics.*

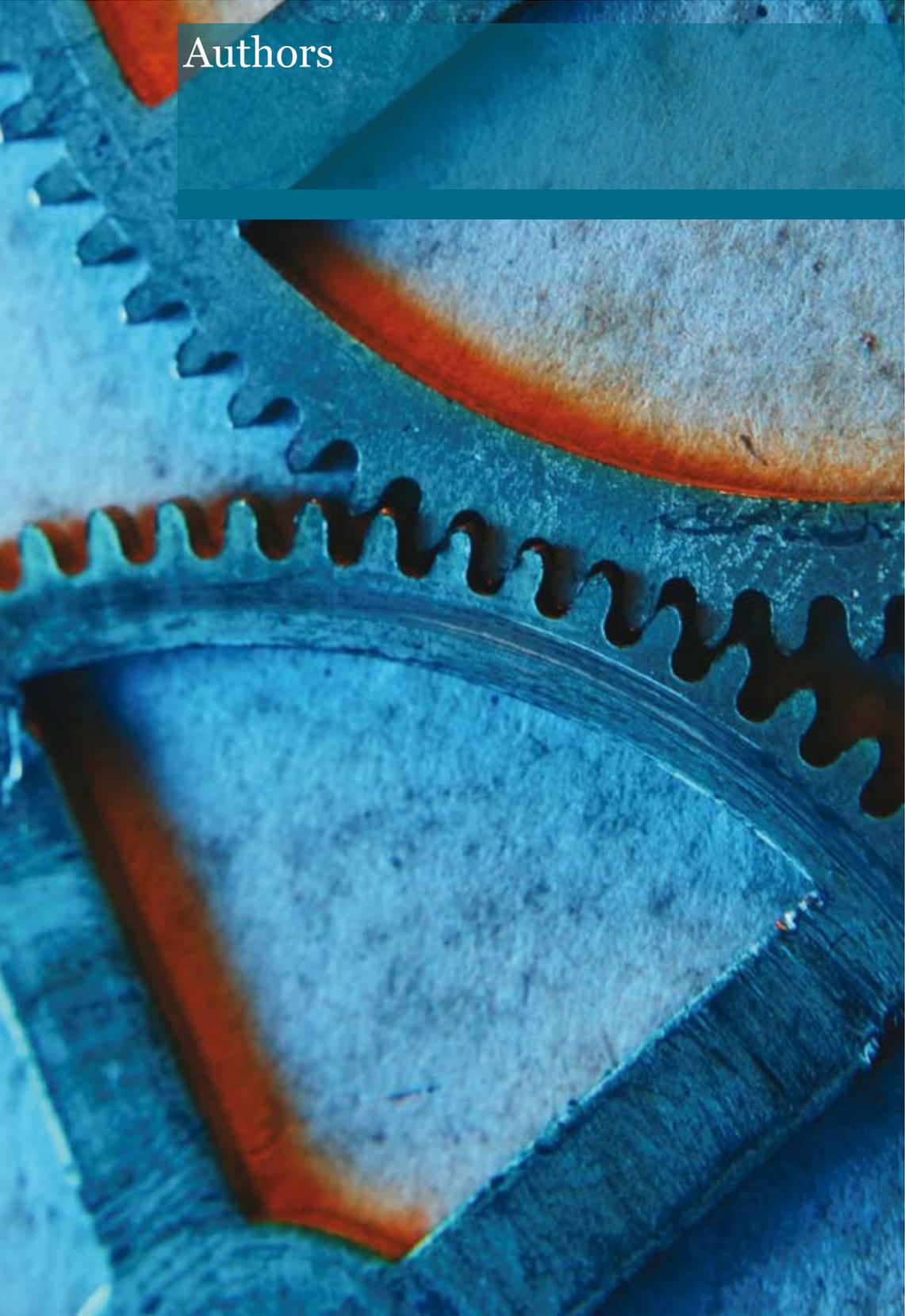
*Note 7. Section 510(k) of the US Food, Drug, and Cosmetics Act requires device manufacturers to notify the FDA of their device at least 90 days in advance of marketing. The review process for this Premarket Notification is more straightforward than Premarket Approval and typically takes less than 90 days.*

*Note 8. These figures are based on historical approval times, and include non-direct review time and direct review time. Premarket approval typically takes around 18 months whereas registration takes some six months.<sup>6</sup>*

*Note 9. The Centers for Medicare and Medicaid Services (CMS) may make national coverage determinations for certain molecular technologies, whereas coverage for most laboratory tests is determined locally by CMS. A local coverage determination is based on a review of current medical practices and clinical data, and procedures for coverage decision are not uniform across localities.*

*Note 10. Reimbursement and billing for molecular diagnostics are performed using current procedural terminology (CPT) codes. Most molecular diagnostics do not have a single unique code assigned. Billing for multivariate tests involves "stacking" multiple codes that describe individual components of the assay. For example, billing for a single Myriad's BRCA panel can involve employing five different codes and 171 separate CPT units.*

- 
- <sup>1</sup> Iakoubova, O. A. *et al.* "Association of the Trp719Arg polymorphism in kinesin-like protein 6 with myocardial infarction and coronary heart disease in 2 prospective trials." *J. Am. Coll. Card.* 51, 435-443 (2008).
- <sup>2</sup> Shiffman, D. A. *et al.* "Kinesin family member 6 variant is associated with coronary heart disease in the women's health study," *J. Am. Coll. Card.* 51, 444-448 (2008).
- <sup>3</sup> Iakoubova, O. A. *et al.* "Polymorphism in KIF6 gene and benefit from statins after acute coronary syndromes." *J. Am. Coll. Card.* 51, 449-455 (2008).
- <sup>4</sup> Zafar, I., Levy, T., & Damergy, S. "Third Wave Technologies Report, Hospital Supplies and Medical Devices," Deutsche Bank, May 12, 2008.
- <sup>5</sup> Davies, L. & Welch, H. G. "Increasing incidence of thyroid cancer in the United States, 1973–2002," *JAMA* 295, 2164–2167 (2006).
- <sup>6</sup> Goodman, C. *et al.* "The Value of Diagnostics, Innovation, Adoption, and Diffusion into Health Care." The Southern California Biomedical Council [online], <http://www.socalbio.org/pdfs/thevalueofdiagnostics.pdf> (2005).



## Authors

The authors of these articles represent a cross-section of McKinsey's global R&D practice. We welcome your thoughts and reactions to the ideas presented. Please email us at [pharma\\_r&d@mckinsey.com](mailto:pharma_r&d@mckinsey.com) or contact the authors individually.<sup>1</sup>

Mark Beards ([mark\\_beads@mckinsey.com](mailto:mark_beads@mckinsey.com)) is a consultant in the London office  
 Eric David, MD ([eric\\_david@mckinsey.com](mailto:eric_david@mckinsey.com)) is an associate principal in the New York office  
 Jerel Davis, PhD ([jerel\\_davis@mckinsey.com](mailto:jerel_davis@mckinsey.com)) is a consultant in the Silicon Valley office  
 Áine Denari ([aine\\_denari@mckinsey.com](mailto:aine_denari@mckinsey.com)) is a consultant in the Dublin office  
 Amar Desai, MD ([amar\\_desai@mckinsey.com](mailto:amar_desai@mckinsey.com)) is a consultant in the Los Angeles office  
 Michael Edwards ([michael\\_edwards@mckinsey.com](mailto:michael_edwards@mckinsey.com)) is an associate principal in the London office  
 Matthias Evers, PhD ([matthias\\_evers@mckinsey.com](mailto:matthias_evers@mckinsey.com)) is a principal in the Hamburg office  
 Edd Fleming, MD ([edd\\_fleming@mckinsey.com](mailto:edd_fleming@mckinsey.com)) is a director in the Silicon valley office  
 Laura Furstenthal, PhD ([laura\\_furstenthal@mckinsey.com](mailto:laura_furstenthal@mckinsey.com)) is a principal in the Silicon valley office  
 Maria Gordian, MD ([maria\\_gordian@mckinsey.com](mailto:maria_gordian@mckinsey.com)) is a principal in the New York office  
 Michele Holcomb, PhD ([michele\\_holcomb@mckinsey.com](mailto:michele_holcomb@mckinsey.com)) is a principal in the New Jersey office  
 Philip Ma, PhD ([philip\\_ma@mckinsey.com](mailto:philip_ma@mckinsey.com)) is a director in the Silicon Valley office  
 Sam Marwaha ([sam\\_marwaha@mckinsey.com](mailto:sam_marwaha@mckinsey.com)) is a director in the New York office  
 Martin Møller ([martin\\_moller@mckinsey.com](mailto:martin_moller@mckinsey.com)) is a principal in the Copenhagen office  
 Troy Norris ([troy\\_norris@mckinsey.com](mailto:troy_norris@mckinsey.com)) is a consultant in the New York office  
 Richa Pande, MBBS ([richa\\_pande@mckinsey.com](mailto:richa_pande@mckinsey.com)) is a consultant in the New Jersey office  
 Paul Rutten ([paul\\_rutten@mckinsey.com](mailto:paul_rutten@mckinsey.com)) is an associate principal in the Amsterdam office  
 Mubasher Sheikh, MD ([mubasher\\_a\\_sheikh@mckinsey.com](mailto:mubasher_a_sheikh@mckinsey.com)) is a principal in the London office  
 Navjot Singh, PhD ([navjot\\_singh@mckinsey.com](mailto:navjot_singh@mckinsey.com)) is a principal in the New York office  
 Saumya Sutaria, MD ([saumya\\_sutaria@mckinsey.com](mailto:saumya_sutaria@mckinsey.com)) is a principal in the Silicon Valley office  
 Tony Tramontin, PhD ([tony\\_tramontin@mckinsey.com](mailto:tony_tramontin@mckinsey.com)) is an associate principal in the New York office  
 Rodney Zimmel, PhD ([rodney\\_zimmel@mckinsey.com](mailto:rodney_zimmel@mckinsey.com)) is a director in the New York office

McKinsey & Company is a global consultancy firm that helps leading corporations and organizations to make distinctive, lasting, and substantial improvements to their performance. McKinsey advises companies on strategic, operational, organizational, and technological issues. Of our 8,000 consultants worldwide, more than 1,600 focus on health care, and 30 percent of these hold PhD or MD degrees. In the past four years, we have completed almost 4,000 engagements in health care, spanning all major sectors and functions of the industry.

<sup>1</sup> This list includes current McKinsey personnel only.

