

# Failure rates in drug discovery and development: *will we ever get any better?*

Drug discovery is an expensive, slow and risky business. An analysis which takes into account projects which neither succeed nor fail suggests that it costs more than \$1 billion to launch a technically successful drug, and takes 12.5 years. Much of the blame for this has been laid at the door of scientific, technical or medical failure of programmes. But more detailed analysis suggests that as much blame should lie at the door of managers unable to make decisions about borderline projects. Implementing a ruthless 'success or die' policy could half the cost and the time to get a drug to market. This may be why 'biotechnology' companies are relatively successful: the Darwinian logic of under-capitalised, over-pressured start-ups means that 'success or die' is the criteria by which managers as well as projects live. However, they are not likely to be inherently faster or cheaper to develop small molecule NCEs than conventional pharmaceutical companies, and so should focus on technology platforms, support services, or genuinely radical medical approaches.

**I**t is widely accepted that the process of creating new medicines is not efficient. The most commonly cited estimate suggests that it costs \$800 million R&D expenditure to launch one new drug, most of the cost being sunk into failures<sup>1</sup>. Perhaps only the movie industry tolerates such enormous costs and failure rates in the search for blockbusters, but they can go from concept to

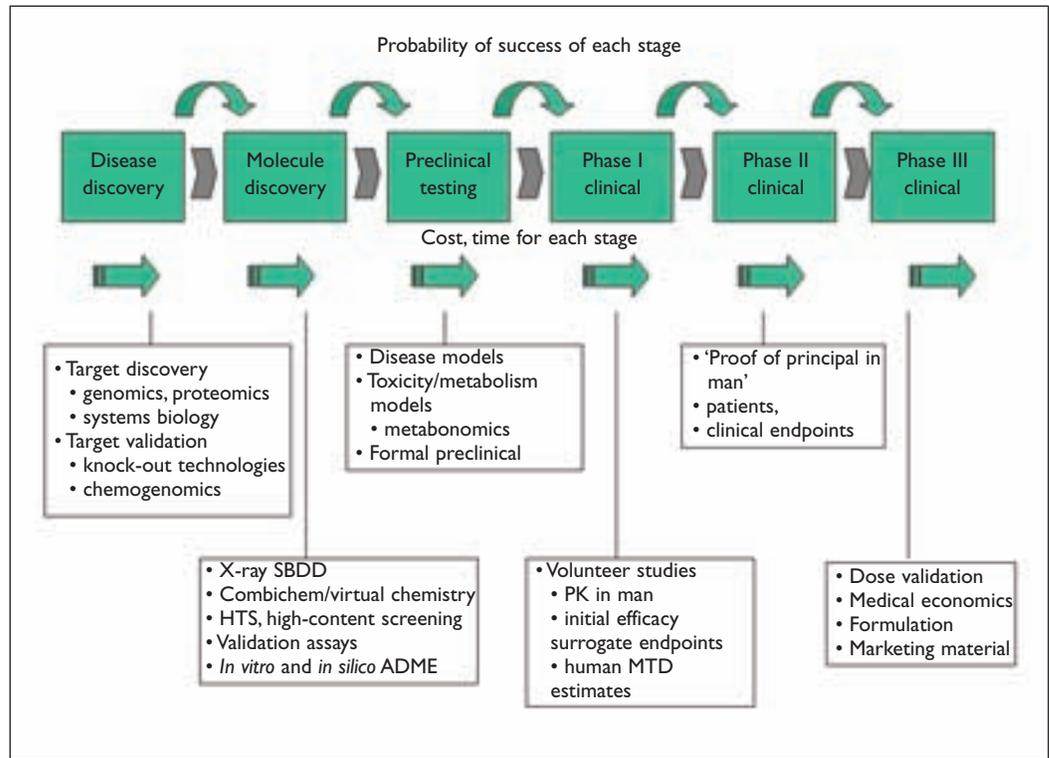
commercial success in a few years, whereas the average for pharmaceutical development is more than a decade.

The \$800 million figure was derived by Tufts<sup>1</sup> from a detailed questionnaire of pharmaceutical companies as to how much each of the stages of drug discovery and development costs, how often it succeeded, and hence calculating the expected

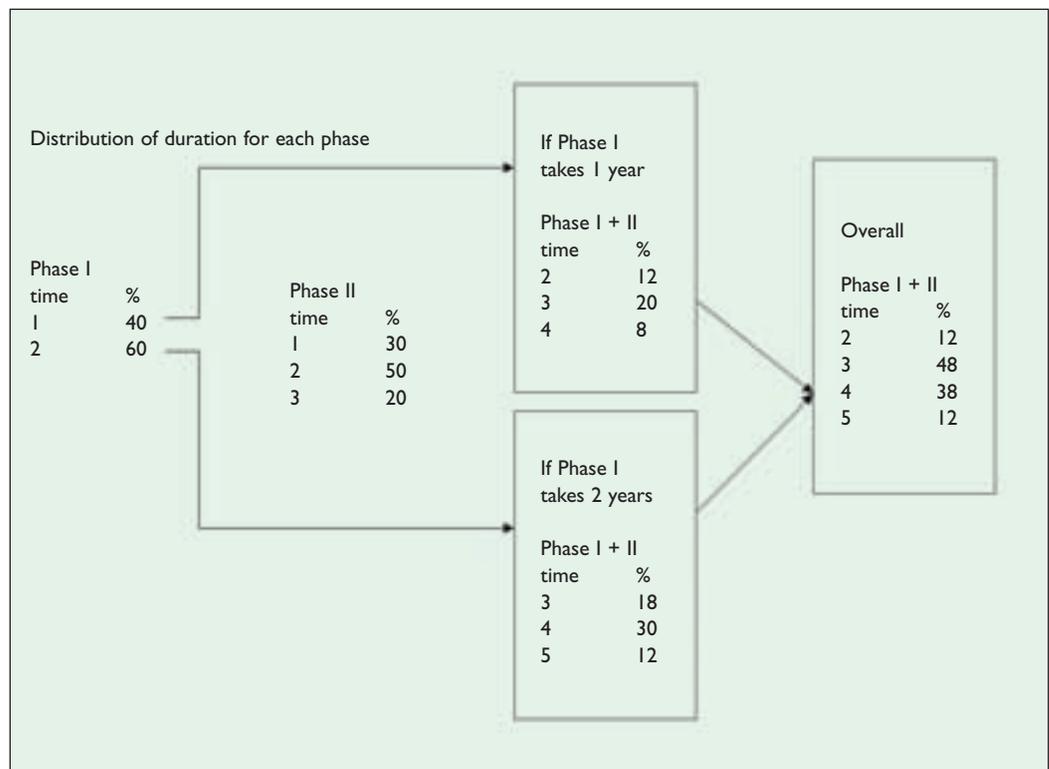
**By William Bains**

---

**Figure 1**  
Cartoon of the conventional drug discovery process



**Figure 2**  
Illustration of method for adding distributions of durations of each stage of a process to obtain the duration distribution for the whole process. Shows are simplified distributions for two stages, and a cumulative distribution of duration for both



cost (and timescale) necessary to produce one successful drug. This approach has three flaws:

- It assumes that the costs and times of each of the steps are independent.
- It assumes that each project is a success or a failure.
- It assumes that pharmaceutical companies know what their costs and failure rates really are (and even if they do know them, are willing to tell someone else).

None of these assumptions is likely to be correct. The complexities of the first point I will ignore, but the second and third could, and should, be modelled, as they illustrate the real ‘value points’ in the pharmaceutical discovery and development chain, and hint at why the biotechnology industry has such difficulty realising the enormous value that it believes it creates.

**How drugs are discovered**

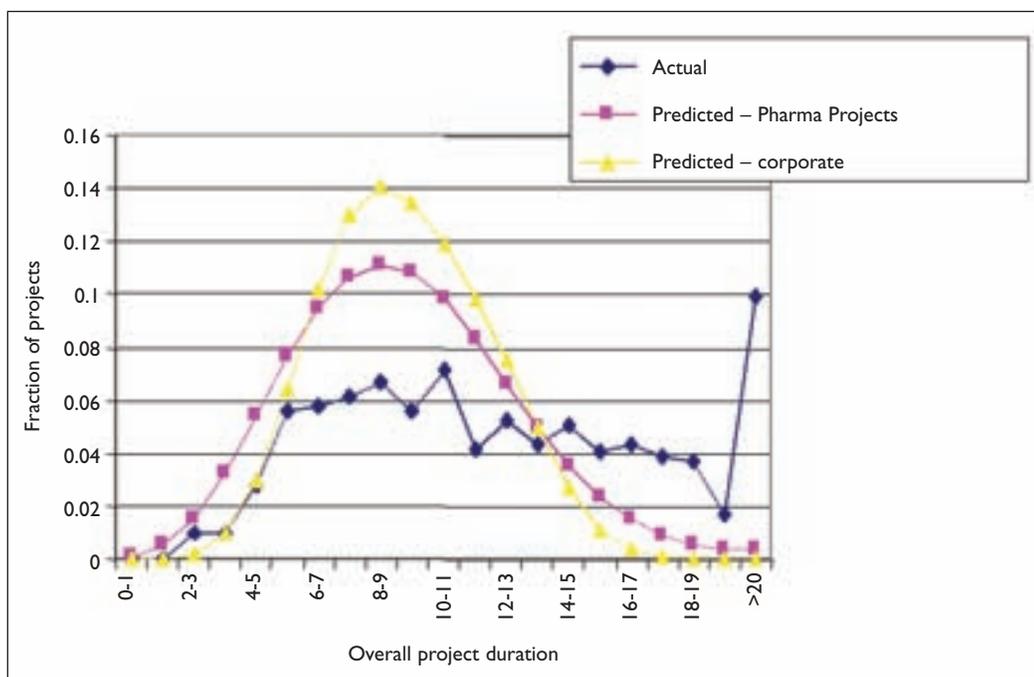
The ‘standard model’ of drug discovery and development (DDD), as presented as slide 2 or 3 of about half of the industry presentations since 1997, is illustrated in Figure 1. The time needed to perform each of these steps can be gathered from ‘experimental’ data from three sources:

- Commercial databases (such as Pharma Projects<sup>2</sup>) which track the development of a large number of specific drugs.

- Company websites, which often list press announcements of commencement and conclusions of clinical trials.
- Conference proceedings, where the same data is given more expansively but often less formally.

The advantage of databases is that they list failures as well as successes – most company websites and presentations do not. The advantage of websites is that it is relatively easy (though time consuming) to track individual projects. The advantage of conferences is that it is possible to find the ‘what might happen’ speculations of companies with new technology. Although the fashion for statements of the form “our chemagenotypic screening can reduce drug development times by up to 30%” faded with the collapse of genomics company stocks after mid-2000, there is a wealth of such data available in the conference proceedings of the larger conference organisers such as IBC<sup>3</sup>, IIR<sup>4</sup> and CHI<sup>5</sup>.

Table 1 lists a compilation drawn from the company papers and presentations on what they reasonably expect to be able to achieve by 2005. It is figures such as these – compilations of data from a wide number of authoritative sources about the cost, duration and chances of success – that are typically used to model the process overall. This compilation is slightly more optimistic than Tufts’ (an overall cost of \$726 million to make one successful drug, including the cost of failures). But the



**Figure 3**  
Cumulative duration of drug discovery and development projects, from first patent date to launch based on a) actual data of the first filed patent date on newly launched products, and the sum of the times takes to traverse each stage of the project as taken from b) Pharma Projects and c) corporate presentations

**Table 1:** ‘Established’ costs, average times and failure rates for stages in the drug discovery process

| STAGE                                 | CONCEPT – TARGET | TARGET – VALID TARGET | VALID TARGET – SCREEN HIT | 2ND ASSAYS | IN VIVO TEST | FORMAL PRECLINICAL |     |
|---------------------------------------|------------------|-----------------------|---------------------------|------------|--------------|--------------------|-----|
| p(failure) (fractional)               | 0.12             | 0.28                  | 0.17                      | 0.10       | 0.08         | 0.34               | 0.2 |
| Cost (\$m)                            | 1.70             | 16.96                 | 14.75                     | 7.89       | 8.03         | 6.35               | 18  |
| Time (years)                          | 0.56             | 0.56                  | 0.68                      | 0.66       | 0.62         | 0.39               | 1.7 |
| Number of projects to launch one drug | 7.91             | 6.93                  | 5.00                      | 4.13       | 3.74         | 3.42               | 2.2 |
| Cost of those projects (\$m)          | 13.42            | 117.48                | 73.67                     | 32.58      | 29.99        | 21.73              | 41  |

qualitative answer is the same – more than a decade and more than half a billion dollars to launch one new molecule. The times and failure rates for actual drug discovery projects as listed in Pharma Projects are listed in **Table 2**, and are broadly comparable.

However, the data do not square with the duration individual of clinical trials, which are also shown as averages in **Table 2**, drawn from individual press announcements of 71 trials from 18 companies. (It is not practical to gather trial failure data from company announcements – reading many company websites literally would lead the browser to believe that many major biotechs had never had a failure.)

The data also do not fit with the ranges of values observed. All these data sources provide a range of values, for which **Tables 1** and **2** only give a summary. From the distributions, it is possible to calculate the expected distribution of the overall time from starting a discovery project to launching the drug, as is done for averages in **Table 1**: a simplified illustration of this approach is shown in **Figure 2**. The result is shown in **Figure 3** for time and failure data taken from corporate presentations and from Pharma Projects. For the ‘average’ project this does not do too badly (hence the relative robustness of conventional predictions of the ‘average’ cost and time to launch a drug), but it substantially underestimates the number that take

**Table 2**

|                         | INDUSTRY EXPERTISE |     |              | PHARMA PROJECTS |              | COMPANY DATA |
|-------------------------|--------------------|-----|--------------|-----------------|--------------|--------------|
|                         | Time (years)       |     | Failure rate | Time (years)    | Failure rate | Time (years) |
| Discovery (all stages)  | 3.1                | 3.5 | 0.57         | 4.03            | 0.69         |              |
| Preclinical development | 0.39               |     | 0.34         |                 |              |              |
| Phase I                 | 1.77               |     | 0.22         | 1.65            | 0.75         | 1.1          |
| Phase II (all)          | 3.5                |     | 0.30         | 2.72            | 0.35         | 1.3          |
| Phase III               | 1.77               |     | 0.1          | 1.86            | 0.58         | 1.75         |

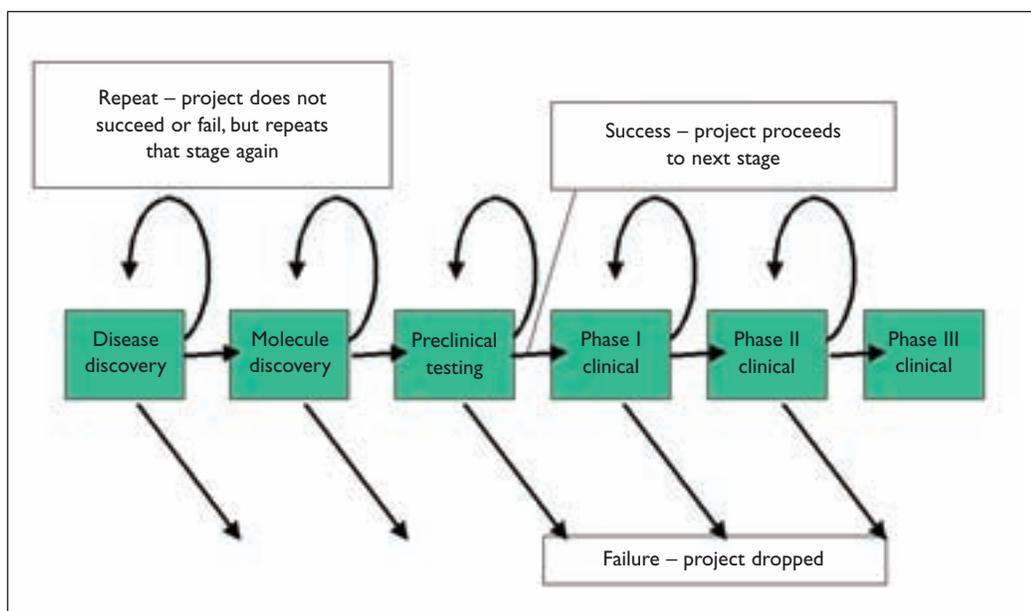
| PHASE I | PHASE IIA | PHASE IIB | PHASE III | REGISTRATION – APPROVAL | PROFIT | CUMULATIVE |
|---------|-----------|-----------|-----------|-------------------------|--------|------------|
| 0.22    | 0.16      | 0.17      | 0.10      | 0.10                    | 0.00   | 0.87       |
| 18.47   | 30.17     | 34.51     | 198.93    | 41.93                   | 0.05   | 379.7      |
| 1.77    | 1.77      | 1.77      | 1.77      | 2.00                    | 0.00   | 12.5       |
| 2.26    | 1.76      | 1.48      | 1.23      | 1.11                    | 1.00   |            |
| 41.66   | 53.21     | 51.16     | 245.18    | 46.48                   |        | 726.56     |

a greater than average time to get from initial patent to product, especially those that take more than 20 years (‘Methuselah projects’).

**The iterative development model**

The reason for this is that the standard model (Figure 1) is flawed – while industry observers are no doubt quite accurate and honest in their assessment of how likely a project is to fail at each stage,

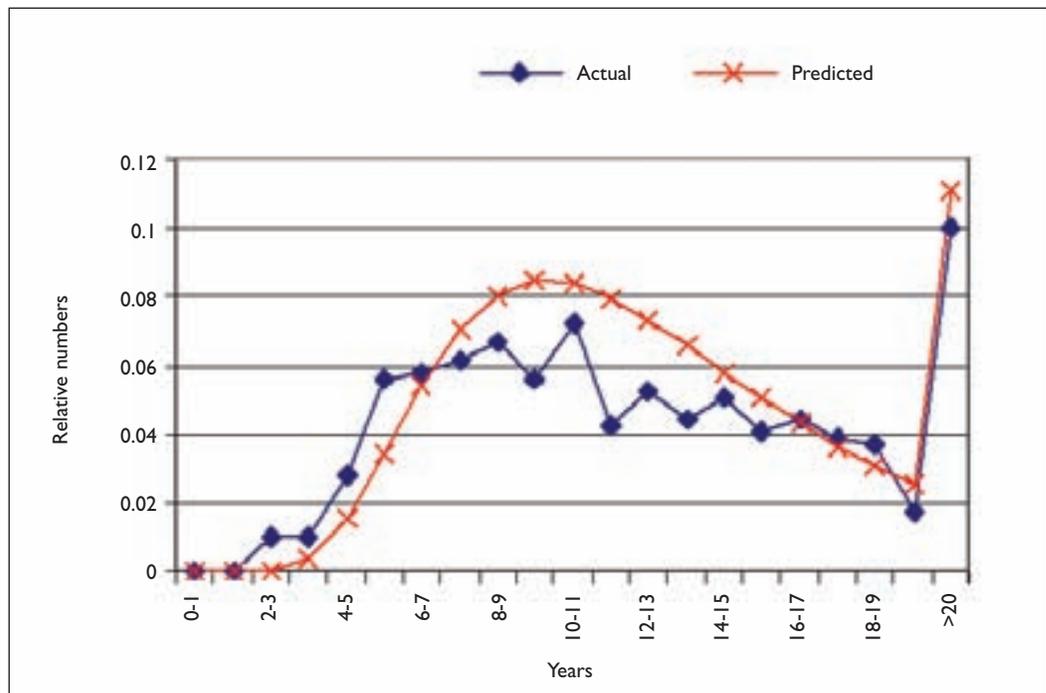
‘not failing’ does not mean that a project succeeds. A project can ‘not succeed’ and ‘not fail’, but instead go round and round an apparently endless series of refinement of dosing, indication, formulation and (within the last decade) reanimation by merger or spin-out. Such projects are not counted accurately in a ‘has it failed or has it succeeded’ analyses. The real process is therefore more like Figure 4. A Phase II trial often produces results



**Figure 4**  
Cartoon of more realistic project flow through the drug discovery and development process

**Figure 5**

Actual duration of drug discovery and development projects from first patent filing to launch, and projected cumulative development time from the model in Figure 4



that point to an interesting effect but which need a further ‘proof of concept’ (ie Phase II) trial. This can be a new dosage or delivery route of the same molecule for essentially the same indication, or it may be for a completely different indication. In short, drug discovery is iterative, not linear. (The reality is even more complex, as clinical results can spawn new research projects and so on: I have not modelled this here.)

So I modelled the process in Figure 4. Each stage does not take ‘a’ time, but rather has a distribution of times. The actual distribution of the durations of each ‘box’ was taken from the duration data from company reports of individual trials summarised in Table 2 for clinical trials, and reasonable estimates made for preclinical and discovery stages based on data in Table 1. The probability (F<sub>x</sub>) of failure was estimated from the data

**Table 3**

|   | DISCOVERY | PRECLINICAL | PHASE I | PHASE II | PHASE III | OVERALL |
|---|-----------|-------------|---------|----------|-----------|---------|
| Time per stage  | 3.5       | 0.46        | 1.4     | 2.3      | 5.1       | 12.9    |
| Chance of failure   | 0.4       | 0.35        | 0.22    | 0.3      | 0.1       | 0.984   |
| Progression probability   | 0.5       | 0.5         | 0.4     | 0.4      | 0.15      |         |
| Chance that one project entering this stage will eventually successfully leave it | 0.58      | 0.52        | 0.42    | 0.52     | 0.44      |         |
| Number of projects needed to achieve one successful launch                        | 30        | 13          | 6.2     | 3.6      | 1.7       |         |

**Table 4**

|                             | <b>DISCOVERY</b> | <b>PRECLINICAL</b> | <b>PHASE I</b> | <b>PHASE II</b> | <b>PHASE III</b> | <b>TOTAL</b> |
|-----------------------------|------------------|--------------------|----------------|-----------------|------------------|--------------|
| Cost per project            | 20               | 3                  | 9              | 93.2            | 59.4             |              |
| Number of projects          | 30               | 13                 | 6.2            | 3.66            | 1.77             |              |
| Cost per successful project | 609              | 39                 | 56             | 341             | 105              | 1150         |

sources for Table 1. The probability of successful progression (Px) was estimated by least squares fitting of the calculated time to move from one phase to another and the observed time lapse between doing so on a project by project basis in Pharma Projects.

The overall result using the ‘status quo’ parameters is illustrated in Figure 5, with a numerical summary in Table 3. This matches the observed duration of projects much more satisfactorily.

Table 3 does not list costs. Putting the costs from Table 1 into the project numbers of Table 4 results in an estimate of \$2.6 billion to successfully launch one drug! However this is unrealistic – most of the estimates in Table 1 will be concocted by taking the overall cost to a company of programmes in each phase and dividing them by the number of programmes – this has the cost of programmes that are repeated built in, and so double counts programmes that are repeated. To compensate for this, the costs in Table 1 can be multiplied by the progression probabilities from Table 3, giving a ‘per step’ cost, which is listed in Table 4. Even with these reduced costs (which are more in line with my experience of actual quotations for different activities, at least at the early stages) the overall cost to produce one successful drug is more than \$1 billion.

Estimating actual cost is not my primary objective, however, but rather discovering how we improve things. The model represented by Figures 4 and 5 and whose output is summarised in Tables 3 and 4 is a more realistic basis for doing this than the ‘traditional’ one in Figure 1: a) it replicates the real time distribution of both individual parts of the discovery process and its overall output, b) it replicates the observed failure and success rates of those parts, c) and it explains the high cost of apparently relatively low cost activities (such as a formal preclinical series) in the context of failures to ‘progress or kill’ a many projects.

**How to improve the process**

Using this model, the sensitivity of the overall cost and time to changes in the process can be mapped: this is done in Figure 6. This analysis suggests:

- As many others have commented, the largest impact on the timescale of drug discovery is to be made in late clinical trials. Changes in any of the earlier stages have relatively little impact.
- That improving the success rate in most phases will actually increase the time it takes to discover a drug, unless it goes along with a parallel improvement in the ability to terminate those projects that do not succeed, rather than allowing them to be taken forward to ‘just one more experiment’. This is very discouraging for the application of new technology in drug development, as this will not merely fail to improve matters in itself, but will provide yet another opportunity for ‘just one more experiment’.
- Costs can be reduced by any improvement in the process: in general, improvements in the success rate are slightly more effective than improvements in the ability to ‘kill’ failures. The exception is Phase III clinical trials. As would be expected, the huge costs of late stage clinical trials means that cost savings are most likely here.
- The only way to reduce costs and reduce time to launch is by reducing repeats, at any stage.
- Almost exactly the same pattern applies even if the NPV of money invested in research is considered – reducing repeats at Phase III reduces the need for so many projects, which reduces the cost of discovery, which has the highest NPV/dollar spent from the viewpoint of the end of the process.

This analysis suggests that, if new technologies such as genomics reduce the failure rate in discovery, this will paradoxically also increase the time taken, although not by a substantial amount. This is quite independent of the increase in cost and time that McKinsey suggested will be the result of

**Figure 6**

Sensitivity of model to 50% improvement in repeat rate or success rate at each stage. Positive values are fractional improvement in the whole process.

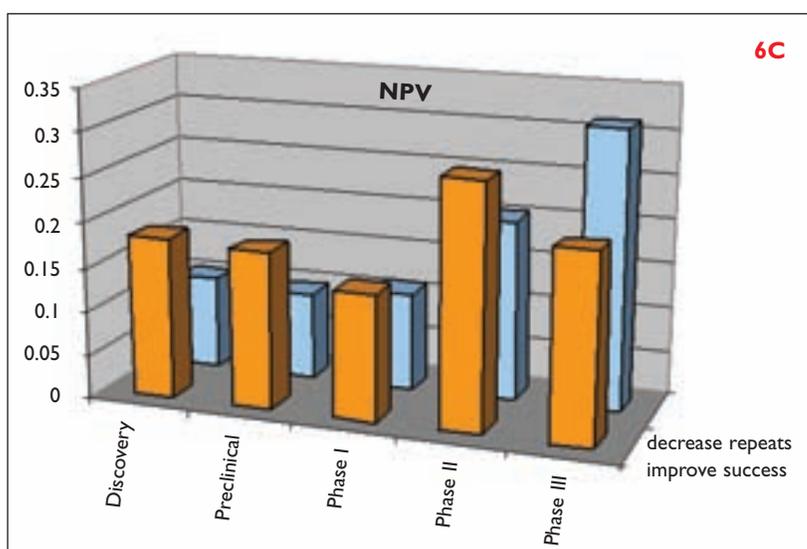
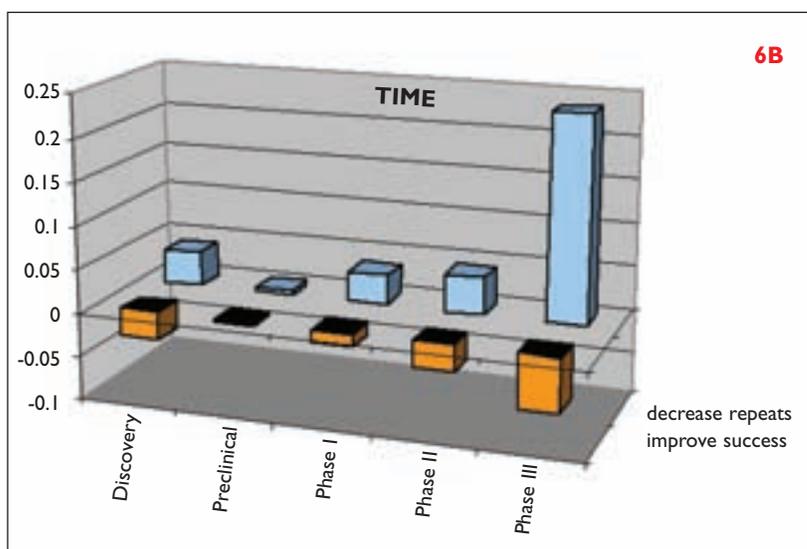
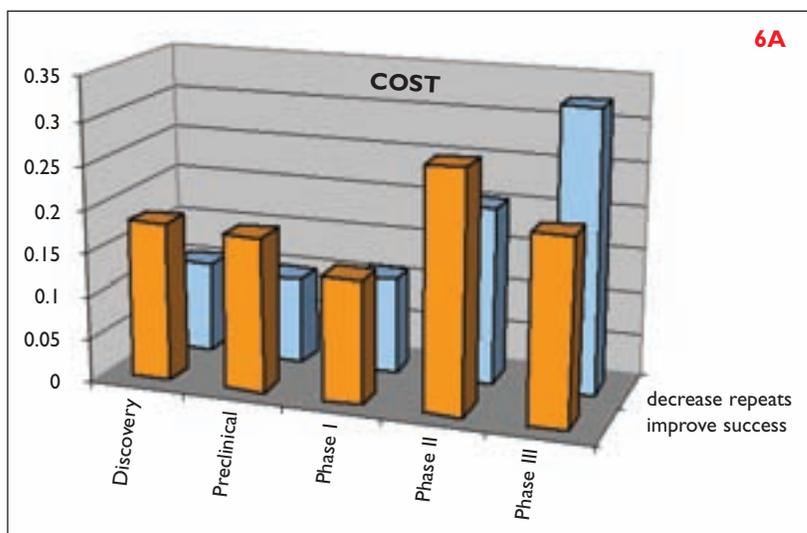
**6A:** cost, **6B:** time, **6C:** Net Present Value cost

genomics over the medium term<sup>6</sup>. They suggest that increase in projects and capacity limitations will have this effect – this model suggests that even if these effects do not occur, and that pharmaceutical companies have infinite capacity to take on new, genomics-derived projects, times for delivering new drugs will still go up with new technology.

The most general approach that this model suggests is worth pursuing to reduce costs and timescales is to reduce the repetition rate. If all projects were classed as success or failure first time, and dropped or passed to the next stage (Progression probability = 1 for all stages in Table 3) then the average drug would cost \$350 million and take six years from initiating a discovery programme to launch even if the failure rates in Table 1 still applied.

This is quite unrealistic – for many indications a ‘repeat’ of phase II or, more commonly, phase III trials is inevitable. Phase II and III trials are used not only to identify dosing and efficacy, but also to determine the exact labelled indication for the compound, especially in complex, multi-layered diseases like cancer and neuropsychiatric disorders. But the level of repetition reflected in Table 3 does not reflect this alone. A new cancer drug entering phase III usually is tested against many different cancers, but in parallel – this is why Phase III trials take so long and cost so much. The repetition reflected in Table 3 is primarily the more serious event, when the hypothesis as to what the drug does is tested in a series of defined patient groups, whether this is called one or many separate trials, the hypothesis is not supported by the statistics, but new experiments or trials are initiated anyway. The logical thing to do is to count this as a failure. However, too often that is not what happens.

By contrast setting all the failure rates to 0 and leaving the progression probabilities from Table 3 unaffected predicts a 16-year average programme length, although at a cost of ‘only’ \$200 million as the single project is tried again and again at each stage until something is found that works. This is a timescale and cumulative cost typical of many academic programmes – their criteria for success or failure not being commercial, it is reasonable to keep on trying new things with the one ‘active’ they have in hand, both to seek new cures and, failing that, seek new knowledge. But it is no way to run a business.



### References

- 1 Tufts University, Tufts Center for the Study of Drug Development, <http://csdd.tufts.edu/InfoServices/Publications.asp>, 2004.
- 2 PJB Publications, Pharmaprojects, [www.pjbpubs.com/pharmaprojects/index.htm](http://www.pjbpubs.com/pharmaprojects/index.htm), 2004.
- 3 IBC Life Sciences, IBC's Drug Discovery Series, [www.drugdisc.com](http://www.drugdisc.com), 2004.
- 4 IIR Conferences, IIR Conferences, [www.iir-conferences.com](http://www.iir-conferences.com), 2004.
- 5 Cambridge Healthtech Institute, Cambridge Healthtech Institute biotechnical/biomedical conferences, [www.healthtech.com](http://www.healthtech.com), 2004.
- 6 McKinsey and Co and LBGE Research, The Fruits of Genomics. 2001: New York.
- 7 Drews, J. Biotechnology's metamorphosis into a drug discovery industry. *Nature Biotechnology*, 1998. 16 (Bioentrepreneur supplement): p.22-24.
- 8 Moscho, A et al. Deals that make sense. *Nature Biotechnology*, 2000. 18(7): p.719-722.

### Implications for 'biotechnology' companies

What does this mean for the biotechnology industry? Most of the science and technology that the industry has developed has been aimed at one of two goals:

- Adding more knowledge to the drug discovery process, so as to make success more certain (eg genomics, SBDD, combinatorial chemistry).
- Creating entirely new therapeutic approaches (gene therapy, antibody therapy, stem cells).

Conventionally, the former would appear to be a much better bet, as the technology is leveraging 100 years' of pharmacology, rather than having to reinvent the rules from scratch with the new technologies. This is the drive behind the recent demand that technology companies become product companies – if you can crack a key issue in the NCE DDD pipeline, then applying it to generate your own drugs is your best bet for commercial success<sup>7</sup>.

However, this model suggests that, even if you prefer the possibility of vast earnings from a blockbuster drug to the smaller but more certain earnings from selling technology today, using your technology to build a drug product company is not ideal. All a company with an astounding new technology can do is reduce the failure rate (and possibly the inherent time) at one stage in the process in **Figure 4**. This does not affect the other factors at all: far from being able to outstrip your pharmaceutical company competition, you will be essentially operating the same management and business model as companies 1,000 times your size (and 10,000 times your capitalisation).

Reflecting this, the reality has been that a rising percentage of genuinely new therapeutics are protein therapeutics, and the biotechnology industry's record on NCE launch is mediocre. The fact many biotechnology companies that are genuinely commercially successful today, such as Genzyme, Amgen, Genentech, Serono, had protein therapeutics as their first 'big product' is seen as the norm, not a blip on the road to biotech's emergence as an NCE powerhouse. I suggest that new therapeutic approaches is the road that emerging therapeutics companies should concentrate on for the future. At the very least, they should be based on a radically new approach to medicine, even if that is implemented in small molecule NCEs.

This is not to say that excellent businesses cannot be made from new technology. The success of companies such as Affymetrix and Applied

Biosystems shows that making something that people will buy is always a good business. Technology product companies will win if there are buyers for their gadgets, from 3G mobile phones to Gigabase gene chips. But to argue that because the technology is cool you can overcome the statistics in **Table 1** is a fallacy: the value of technology that improves the failure statistics in **Table 1** will be realised (if at all) by selling the technology, not by using that technology to make new drugs.

Since the mid 1990s a few companies have tried to build a new business model, using technology or information to help pharmaceutical companies 'fail fast', notably in the ADME services and prediction area with companies such as Camitro (now part of InPharmatica), Cyprotex and Pharsight. My analysis suggests that this should be enormously valuable, but realising this value depends on the pharmaceutical industry recognising it. Genomics discovery deals valued in the \$100 million range were a 'norm'. To my knowledge, no process optimisation deal with a biotech company has passed the \$5 million mark. This is the fallacy in Moscho et al's analysis of the value chain<sup>8</sup>: in theory one can argue that a product that improves a process and so adds huge value to the pharmaceutical client should itself be hugely valuable, but if the client simply will not pay (because they do not believe the extent or the nature of the problem) then the value will not be realised. I hope this analysis will go a little way to providing more ammunition for those companies seeking to help pharmaceutical companies to get their programmes to fail efficiently.

### Acknowledgements

I am grateful to Liz Holt for not believing a word of an earlier version of this, and to Info To Go Ltd ([www.infotogo.co.uk](http://www.infotogo.co.uk)) for summary data from Pharma Projects.

A more detailed set of Tables of figures for the model presented here are given on [www.rufus-scientific.com](http://www.rufus-scientific.com)

DDW

---

*After an academic career in the UK and USA and consultancy work with PA Consulting Group, William Bains joined a Venture Capital group in 1996 and started his first drug discovery company, Amedis Pharmaceuticals, in 2000. William is now an independent entrepreneur, working with several start-up companies and teaching company formation and financing at Cambridge University. William is author of numerous papers and patents and three books.*