

Learning to improve the decision-making process in research

Andrew Chadwick* and Michaela Hajek, PA Consulting Group, Cambridge Technology Centre, Melbourn, Hertfordshire, UK SG8 6DP; *e-mail: andrew.chadwick@paconsulting.com

'The true worth of an experimenter consists in his pursuing not only what he seeks in his experiment, but also what he did not seek.'
Claude Bernard (1813–1878).

In drug discovery and development, the decision-making process is based on a huge amount of scientific data.

However, given the sequential nature of the overall pharmaceutical R&D process, only information localized to the individual stage is available to promote compounds to the next level. One of the major challenges in R&D is to increase the breadth of information available to the compound selection process, considering not only scientific evaluation criteria but also a forward-looking view of commercial and operational consequences. This requires the reliability of selected tests to be defined, as well as the consideration of the potential value of the overall project.

Thirsty for drug candidates but drowning in a flood of data

The falling productivity of R&D is putting pharmaceutical companies under pressure to improve the effectiveness and efficiency of their discovery methods, which has proven to be difficult because scientists often have to make decisions based on inaccurate and incomplete information. Researchers can be overwhelmed with data about their specialized area and yet find it hard to see the big picture about whether or not to progress a compound, or a series of compounds, through the hits-to-leads-to-drug candidate process (Figure 1).

At the early stages of research, a large number of compounds need to be tested. Typically, the information

available is restricted to high-throughput screening (HTS) investigations and computational predictions. It has been estimated that HTS activities alone have produced a 10-fold increase in network traffic. Many companies are using computational, visualization and data-mining techniques to cope with the flood of data [1]. In our work with pharmaceutical research groups, we are often told that the assimilation of this information, to confirm hits, to identify structural patterns, and to choose the best lead series, is rate limiting. In addition, it can be difficult to determine how many lead series should be assessed in parallel. Potentially, an increase in the number of compounds tested means more choice in the later stages of candidate selection, and perhaps a backup that is nearly as good as the main candidate, but this has to be set against the greater load on the available testing and synthesis capacity. Therefore, most companies start filtering compounds and series early in the discovery process.

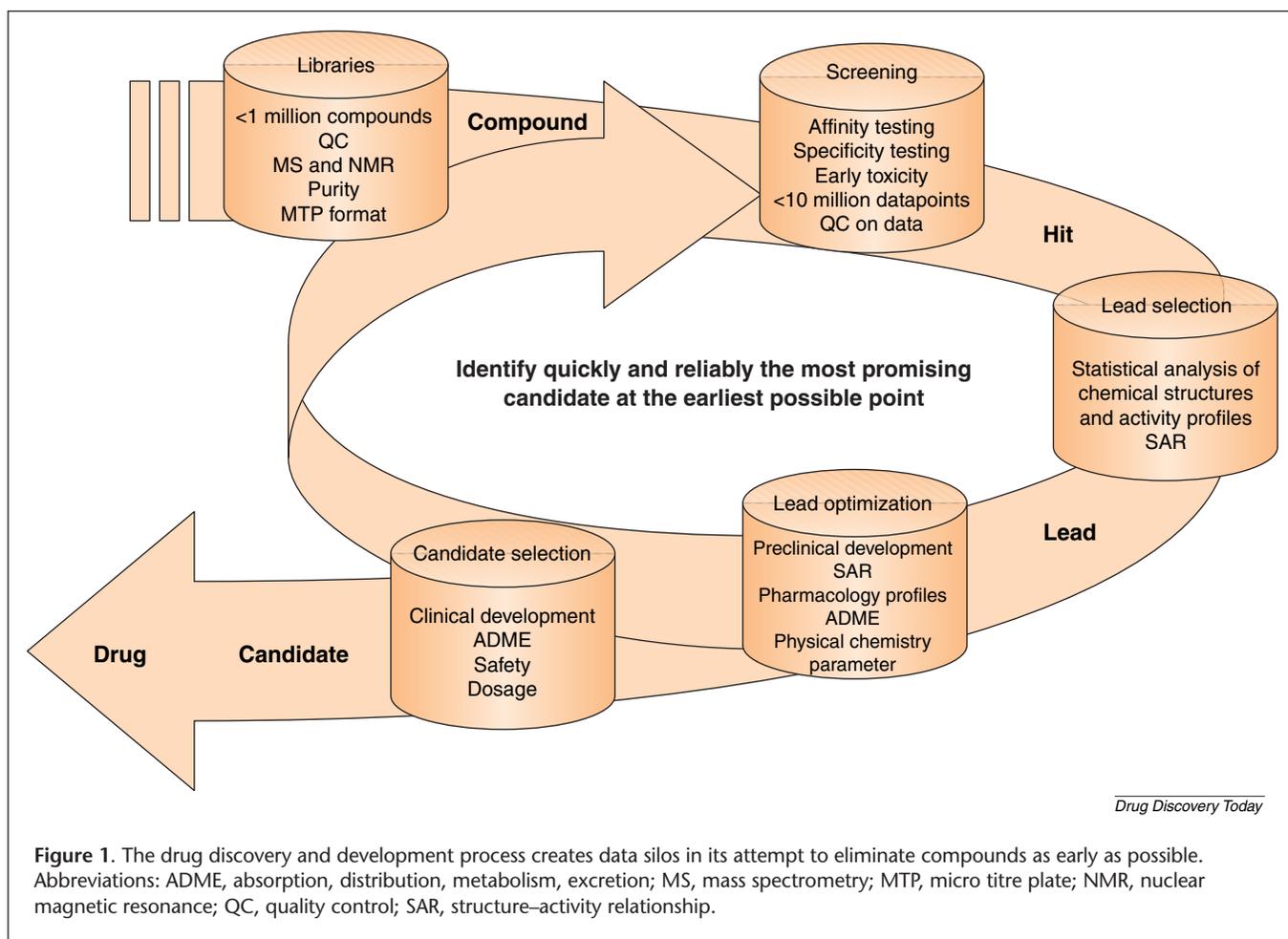
Library pre-selection based on chemical features (virtual screening) accomplishes filtering with regard to recommended physico-chemical properties of the tested compounds. HTS assesses biological activity and perhaps target specificity. However, these early choices typically exclude measurements of absorption, distribution, metabolism, excretion

(ADME) and toxicity. As a consequence, reliable information about important factors, which will determine whether or not a compound or series will make a good drug, is not available to the early compound evaluation process.

Challenges of research planning

A good development candidate must pass reliable tests spanning all the criteria that determine product value. Before the start of human safety and effectiveness trials, the likely behaviour of the compound in humans has to be inferred. Errors in this assessment lead to missed opportunities and to the majority of development failures, with the most frequent causes being pharmacokinetic (PK) difficulties (39%), lack of efficacy in humans (30%), animal toxicity (11%) and adverse effects in humans (10%) [2]. Therefore, research faces a hard challenge: how to use its resources and results to more accurately predict outcomes in humans.

Although there is a move towards the use of ADME or toxicology predictions in parallel to HTS to predict the 'drugability' of the hits [3], this approach is still controversial and might not give the hoped-for benefits [4]. The application of these approaches to library or screening set selection could reduce the diversity of the potential leads. Furthermore, the criteria important for candidate selection are not necessarily all relevant to the initial search for a pharmacophore [5].



To get the best possible end-to-end R&D performance, researchers have to critically assess their way of working and to challenge the reliability (predictive performance) of their screening methods and the sequence of combining their different tests.

Reliable early predictions for bioavailability and PK criteria are hard to obtain because *in vitro* ADME tests each cover only part of the complexity of *in vivo* models [6]. The development of highly reliable *in silico* tools that predict specific ADME parameters from chemical structure alone will take years because large and diverse training sets are needed. In particular, the calculation of ADME parameters of novel chemical compound classes creates difficulties for prediction models. As new families of compounds are developed, information from previous libraries has to be

updated to ensure relevance of the predictive model. Therefore, in seeking 'early attrition' to minimize expensive late failure there is the risk of rejecting perfectly good compounds [7]. Research project managers and resource planners need to solve a difficult planning problem – how do they organize the sequence of predictions and tests, and set cut-offs on these criteria, to get the highest possible output rate of quality candidates from the available resources? This requires scientific information, as well as a forward-looking view of product value and cost and also risk assessment of future R&D [8].

How, in practice, can research groups plan in a way that combines the scientific evidence about different candidate criteria with business and operational considerations, including expected value, resources and capacity (Figure 2)?

Good plans combine scientific and business thinking

Many companies seek to agree 'best practice' project processes and metrics. This is a difficult balancing act because even after target validation, the aims of projects differ according to the nature of the therapeutic drug profile, the available assay technology, and the state of advancement and competition in the relevant therapy and market. The possible requirement for the drug to cross the blood–brain barrier, and the relative tolerance of side-effects in most present-day cancer treatments, are obvious examples.

The methods that a project uses, the order in which they are applied and the best cutoffs all depend on method reliability. Reliability has two important aspects: false positives and false negatives. Although scientists

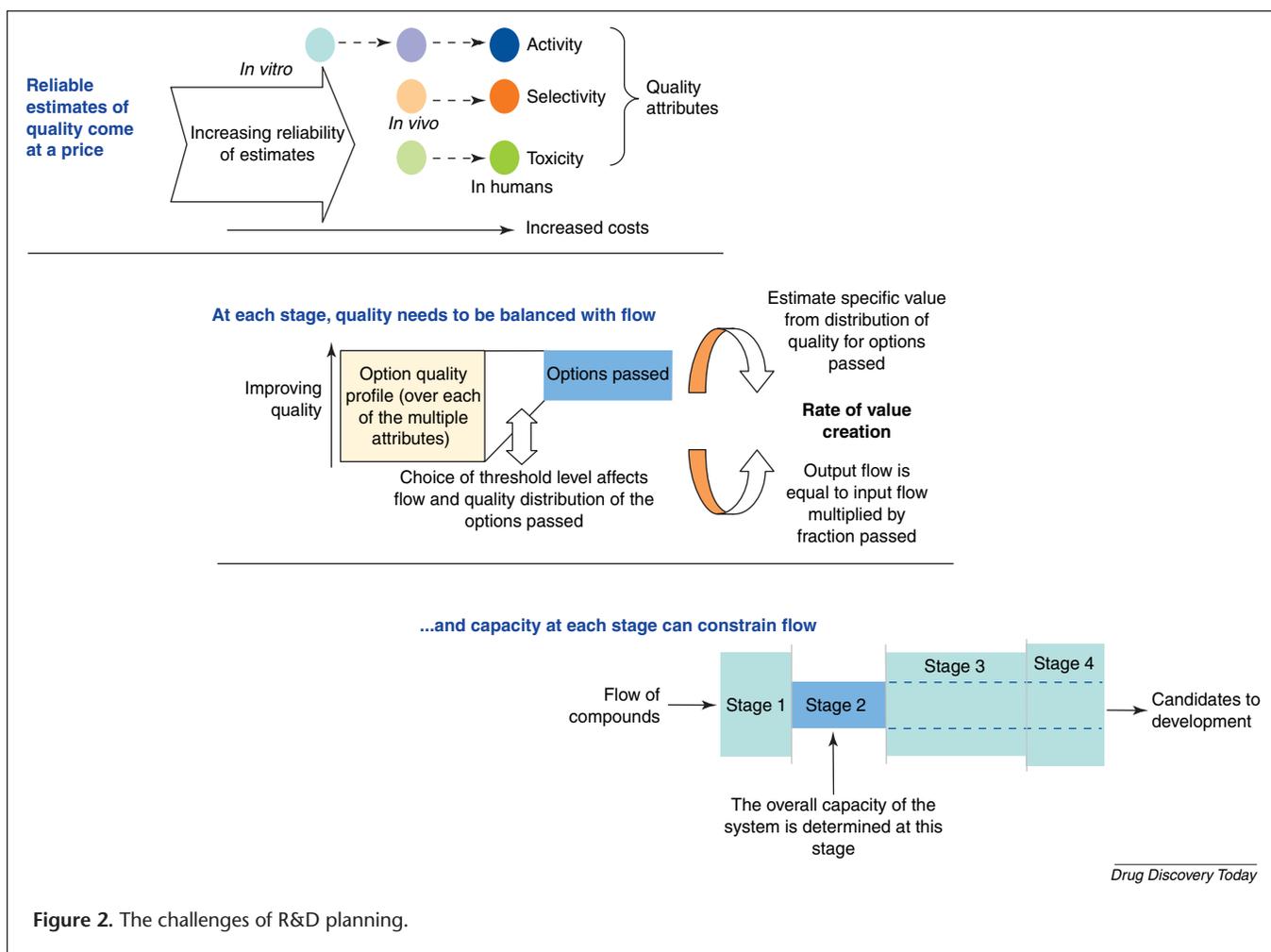


Figure 2. The challenges of R&D planning.

understand and seek to avoid false positives, false negatives have to be taken into account because they can lead to loss of value. The extent of this loss depends on the commercial value of the project and the stage of research. In the earlier stages, it might be easy to find replacement compounds without undue effort. However, in later stages, candidate replacement is more expensive and so projects with a potentially large payoff should be prepared to progress a greater fraction of their options (if they cannot improve method reliability). Therefore, successful planning relies on an understanding of method reliability, of the required quality of development candidates and of the potential value of the drug.

Planning is further complicated by differences between the capabilities of

different sites and organizations. If a site has spare in vivo screening capacity, then more lead series could be followed up in parallel. However, in vivo screening, using a particular test model, only adds value if the chosen parameters contribute sufficiently to the certainty of the predictions concerning outcomes in humans. Thus, a critical appraisal of predictive reliability is just as important for in vivo as for in silico and in vitro methods. Even when available, reliable predictions tend to be slow and costly.

Companies can improve research productivity by planning end-to-end

Companies can only get the best out of research by combining scientific, business and operational considerations. They are already good at interpreting

scientific results at the individual stages of the compound progression process. However, companies might need to get better at building business and operational aspects into the stage gate decisions. In addition, in larger companies in particular, it can be hard to share resources between sites and therapeutic areas because of communication barriers. People need to see and understand the possible benefits before committing to change. To help achieve this, planners and resource managers need a high-level simulation of end-to-end research that can be tailored to account for therapeutic area and project differences and that shows the potential value and risk of a course of action. This model will be founded on assumptions about clinical and market needs and risks.

These assumptions have to be agreed close to the launch of a research project. Such simulations start from a scientific model of cause and effect, build in estimates of test method reliability and then add information on the major variable costs and the 'net value' for a development candidate with a given profile.

We envisage two main applications:

- Project and technology planners will be able to assess a proposed sequence of compound evaluation tests and compare it with alternative routes, with a shared view of potential downstream consequences built into the model.
- Resource planners who face the problem of eliminating a bottleneck will be better placed to understand the pattern of throughput and attrition across the research stages, thereby avoiding the creation of a new bottleneck elsewhere. Their strategic goal should eventually be to create a 'balanced pipeline' of research in which each stage maximizes value from available capacity allowing for year-on-year fluctuations in throughput. For this application, the simulation model must incorporate relevant investment costs for new capacity.

In this type of simulation, the details of individual test timing and of logistics in material management can be neglected; the emphasis will be on the balance of flow and quality stage-by-stage across research and how this would be affected by different methods and cutoffs. The impact of sharing resources can be explored, first between projects, and then, to a limited extent, between therapeutic areas that might have unique methods and dedicated resources. This will require multiple variants of the basic model. The limiting resource at a given time within a research facility might sometimes be laboratory facilities or sometimes specialists appropriate to the stage of work.

Such a simulation approach directly addresses the strategic need for the discovery organization to maximize the value that it adds. The joint goal for improvement is simultaneously to increase throughput and to deliver candidates of high 'quality', while containing the costs of operations. For a development candidate, quality means high potential market value, combined with a relatively low risk of failure in development. The common denominator here is economic value. It is already common practice to put numbers on the value of development candidates, for example, within a portfolio of projects or for licensing. However, research groups can also benefit from a quantitative model for how their decision processes influence the chances of a research project leading to a given 'quality', and hence value, of the candidate that it passes on into development.

By exploring the potential for an improved decision-making process, research managers will not only have a greater understanding of the root causes of certain events (e.g. late-stage failure), but will also perceive the potential impact of certain actions (e.g. investment into early ADME tools) on their projects. In addition, this approach opens doors for implementing an active learning [9] process that would enable researchers to make more-informed decisions in the future by combining their existing insights with new empirical evidence as it becomes available, and simultaneously reviewing and actively seeking to fill gaps in the evidence. This evidence could include information acquired from passed and failed targets, compounds, leads and drug candidates, and will contribute to a continuing re-evaluation of the discovery decision-making process.

Scientists are already using simulation and learning tools in the scientific domain

In the early stages of research, scientists are already using quantitative analytical

approaches, for example, data mining and pattern recognition, to cope with the volumes of data available. There is increasing acceptance of neural networks or belief (Bayesian) networks, which recognize and retain patterns from data and make 'hidden' information more explicit. These methods are used for classifying and clustering of compounds and their practical relevance has been shown, for example, in generating screening subsets of compound libraries representing the chemical diversity of the entire library [10] and in the clustering of drugs and non-drugs [11].

Adding in business and operational factors will improve productivity

In an approach we call ARBITER (Architecture for Reliable Business Improvement and Technology Evaluation in Research), we have recently extended artificial intelligence (AI) and statistical approaches beyond analysis of scientific evidence to wider use in research planning. This approach combines a multivariate view of compound 'quality' with capacity loading and expected value information. By interactively simulating the impacts of different test sequences and cut-offs, the user can understand not only the types of compound that will be progressed, but also the implications for overall success and attrition rates and the potential location of bottlenecks that could need more resource or automation.

This is a step towards a fully value-based approach to research planning. The potential benefits arise from a mix of raised throughput of candidates and a higher average of the standard of drug candidates. This improvement is started by optimizing the use of current information about reliability of methods, and then sustained through systematic learning from new evidence (continuous improvement, or 'closed loop learning').

In contrast to the currently available commercial IT tools that help researchers to pick a winning compound [12], ARBITER helps research managers to pick the approach to research that has the best chance of creating winners, without having to eliminate too many losers at great expense in later stages of R&D. The design of ARBITER (Box 1) enables it to address the following issues: the best screening sequence; the required level of accuracy of predictions to warrant the choice of method at a given stage of research; the effect of a correlation between test parameters; the dependence of method reliability on broad aspects of the 'chemistry' or the target.

This approach will help the user to determine how best to build specific chemical and biological assay sequences for individual projects that are best suited to their preferred areas of chemistry.

Status of ARBITER

To explore this approach, we have constructed and populated a prototype simulation system that allows for differences between companies in the definition of and relative capacities of research stages. Candidate values are based on product value variations between and within therapeutic areas, and on the typical yields and costs of a development stage (based on our broad industry experience and a range of published surveys). The base model of cause and effect includes public-domain information on distributions of important chemical parameters and on the reliability of some of the main research methods. This prototype was intended as a proof of concept, and the knowledge base will need to be customized to the assays and library characteristics of a specific organization. However, we can already demonstrate how the rate of value added by the discovery function can vary under different working practices, with

surprising conclusions. For example, if a bottleneck exists at a given stage, the use of a prediction to filter out options at an earlier stage can increase the overall throughput of candidates but only if this prediction has a reliability close enough to that of a more traditional method used within or after the bottleneck. If the reliability of a method has not been accurately estimated, then even the best worked-out plan for a test sequence using that method could yield less potential value than a research plan derived after the update of the model of cause and effect (belief network) with new results. Typical differences in value between these scenarios are 10% of the value of discovery output.

Companies must realize that a 'quick fix' approach does not work

There is evidence that technology investments into pharmaceutical research have been oversold in the past because of a failure to recognize that the business problem was a lack of certainty rather than a lack of volume.

The process of reliable business improvement is not a quick fix. It starts by capturing and sharing what people know about how decisions are made. In addition, research managers have to build a culture that enables a shared commitment to an objective review of evidence. Understandably, it is hard for project managers to end their project. Furthermore, it will take time to build up the evidence to support change and to improve understanding of the sources of failure in R&D, which includes an unbiased analysis of the reliability and chance of failure of individual tests and predictions. For example, companies must realize that the elimination of a 'couple of false negatives' per screening round might have a greater business impact than previously thought.

The same approach can be applied to technology evaluation for research. The

payoff from many modern investments in research technology comes from a higher standard of information, rather than from increased raw throughput. Therefore, technology evaluation demands attention to the complex interactions of the decision-making process. Companies apply various predictive tools during the compound selection and lead optimization process, yet find it difficult to describe quantitatively the value that these tools add to the overall decision-making process. ARBITER can help companies to analyse the usefulness of applying predictions at the various stages of research.

Before risking time and resources for technology change, the full consequences of different working practices must be considered – the people costs of change, as well as the capital costs. The use of a simulation approach such as ARBITER can help R&D leaders avoid the following expensive mistakes: spending more money on testing (to anticipate attrition) than would be needed to continue work up to the point of the predicted failure; running the screens and selection decisions in the wrong order (given the costs of different methods and the capacity and resources available); using methods of poor reliability that, if sensitive enough to flag later failure, are not specific enough to keep sufficient viable options in the pipeline; trying to discern patterns of hits when there is too much background noise in the data, rather than improving the quality of assay results or the use of historical information; failing to exploit the multiple data sources that are available in parallel (for example, if a selective compound is desirable then results from several assays should be examined simultaneously); losing flexibility in planning by assuming that one research plan fits all therapeutic areas.

Any support process for decision-making is only as good as the

Box 1. Principles of ARBITER

ARBITER uses probabilistic networks for the statistical modelling of scientific cause-and-effect relationships (to explain attrition), combined with a model of demand on capacity for each stage in the sequence of tests. This demand depends on tests applied in a particular stage and on the thresholds set for tests in previous stages. The new proposed architecture combines three elements as detailed below.

(i) A belief network for evidence-based modelling of cause and effect [14] (which is preferred to a neural network, because Bayesian statistical modelling supports the earliest stages of learning in which only subjective judgments could be available [15]). The nodes in this model represent probability distributions that are based at first on scientific judgments and are then adjusted in a training process. Training can use industry-wide data and, within each user company, in-house results that represent either complete project histories or calibration comparisons between pairs of methods. The links between nodes represent conditional probability (i.e. 'cause-and-effect' relationships) and provide a way of helping scientists to share results and to build a consistent model of influences between their area of specialist interest and wider discovery outcomes. These networks have already been applied to automotive-claims screening, learning from field inspections the characteristics of claims that have the highest probability of being overwritten. Within a model of pharmaceutical research, the nodes can include:

- variables that can be tested or calculated [e.g. binding affinity, permeability, solubility and calculated logP (ClogP)];
- molecular descriptors that might influence the reliability of a prediction or measurement method elsewhere in the model (allowing for methods that are used only within some areas of chemistry, and also for compound subsets to follow differing screening sequences);
- variables that could be of a more subjective nature and are used to score a candidate or to estimate its quality (quantitative factors, such as estimated potency, or binary variables, such as 'mutagenicity', that incorporate probability estimates).

In a model that includes early development stages, information acquired about variables that stand for measured outcomes not known to research (e.g. long-term toxicity) can be used for feedback and refinement of the development candidate scoring.

(ii) Estimates of how future development project value, costs and risks depend on the characteristics of the 'product candidate' that discovery forwards to development.

This information is often available as a result of portfolio valuation work involving research, development and commercial functions.

(iii) A steady-state throughput and attrition model of R&D processes in which the throughput at each stage can be constrained by available capacity. The modelling of available capacity requires the understanding of resources and the extent to which they are shared. Important resources are typically people, equipment and material or animals, but the crucial resource varies between the stage and the level of automation of research. Organizations differ in the extent to which they share resources between projects and therapeutic areas. Therefore, the level of resource analysis within the simulation must be tailored to the specific need. This approach has the potential to deal with year-on-year fluctuations in throughput but has, at root, a flow model balancing the flow of work as a rate (e.g. compounds per year) against the capacity for processing that work. When necessary, we assess this capacity with the help of more-detailed discrete event models. The workload at each stage depends on the output from the previous stage, which is determined by the decision criteria applied at that stage and by the relevant distribution of compound 'quality' within the belief net. This distribution is itself contingent on the criteria applied in all earlier stages, including library selection.

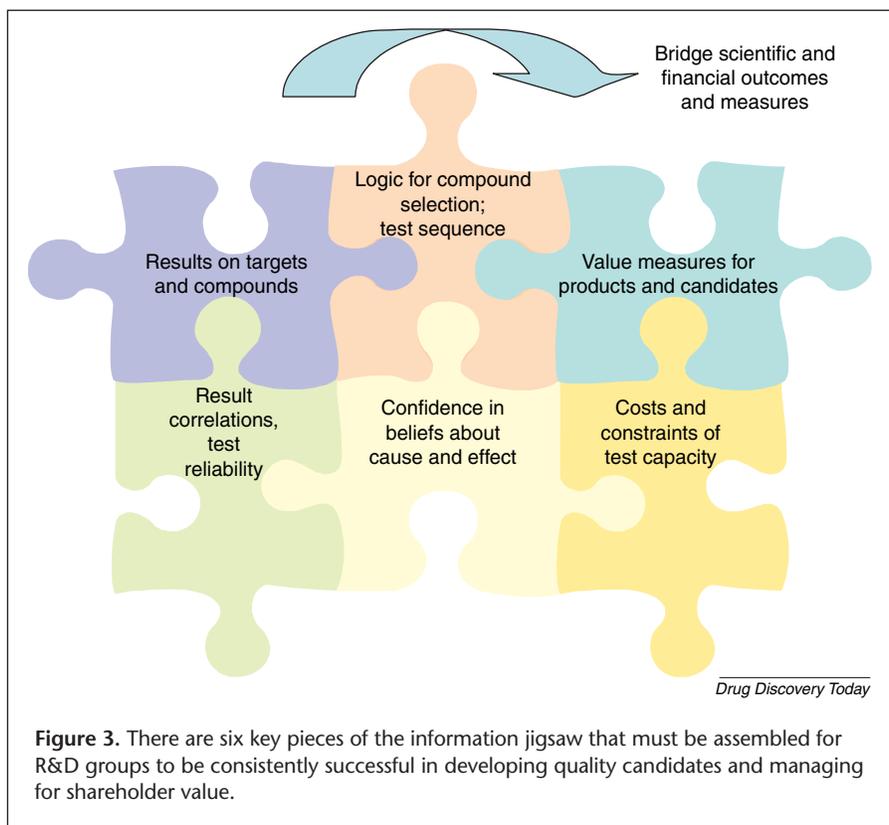
In the simplest case, we compare the economic effectiveness of various alternative, or complementary, methods to estimate a single underlying factor (e.g. *in silico* and *in vitro* findings on mutagenicity as proxies for *in vivo* measures). The use of a belief network allows for the multivariate nature of compound 'quality' and for correlations between factors, but relies on there being some understandable structure in the relationships between 'quality' factors and test findings; this is not a 'black box' model.

For any given set of assumptions about tests and cutoffs, the simulation calculates the rate at which the drug discovery function is adding value, which is based on the average throughput of development candidates and their average value. In addition, variable costs can be introduced; however, from our experience, in the short term discovery groups are looking for the best application of fixed assets and technologies. There is the potential to optimize cutoffs, applications of tests and available capacity, which facilitates the valuation of many research technology investments. Finally, this assessment depends on the confidence in the information acquired by the user, and must be revised after significant advances in method reliability, or revisions to candidate value measures.

information provided. Belief networks (based on Bayes' theorem) are 'a critical advance, because no decision process that must estimate the likelihood of an

uncertain outcome can ever do better than a Bayesian estimate of that probability' [13]. When information is lacking, the exercise of modelling cause

and effect can help to identify important gaps and inconsistencies where new information, which is relevant to key decisions, could make a



significant difference to the effectiveness of the research process and of research planning.

Improved communication between scientists, planners and commercial market analysts, as well as an

architecture that breaks down barriers between scientific, statistical and commercial considerations (Figure 3), will be essential to systematic improvement in the research decision-making process.

References

- 1 Gedeck, P. and Willet, P. (2001) Visual and computational analysis of structure–activity relationships in high-throughput screening data. *Curr. Opin. Chem. Biol.* 5, 389–395
- 2 Kennedy, T. (1997) Managing the drug discovery/development interface. *Drug Discov. Today* 2, 436–444
- 3 Oprea, T.I. (2002) Virtual screening in lead discovery: a view point. *Molecules* 7, 51–62
- 4 Lipinski, C.A. (2003) Chris Lipinski discusses life and chemistry after the rule of five. *Drug Discov. Today* 8, 12–16
- 5 Hann, M.M. *et al.* (2001) Making lead discovery less complex? Sheffield Computational Chemistry Symposium (<http://cisrg.shef.ac.uk/shef2001/confbody.htm>)
- 6 Sietsema, W.K. (1989) The absolute oral bioavailability of selected drugs. *Int. J. Clin. Pharmacol. Ther. Toxicol.* 27, 179–211
- 7 DeWitte, R.S. (2002) On experimental design in drug discovery. *Curr. Drug Discov.* February, 19–22
- 8 Greuel, J.M. (2002) The R&D value conundrum. *Curr. Drug Discov.* July, 37–42
- 9 Raju, G.K. and Cooney, C.L. (1998) Active learning from process data. *AIChE J.* 44, 2199–2211
- 10 Agrafiotis, D.K. and Lobanov, V.S. (2000) Nonlinear mapping networks. *J. Chem. Inf. Comput. Sci.* 40, 1356–1362
- 11 Ajay, A. and Walters, W.P. (1998) Can we learn to distinguish between “drug-like” and “nondrug-like” molecules? *J. Med. Chem.* 41, 3314–3324
- 12 Beresford, A. (2003) *Compound Prioritization for Drug Discovery*. Arqule WO03/042898A2
- 13 Glimcher, P.W. (2003) *Decisions, Uncertainty and the Brain*, MIT Press
- 14 Pearl, J. (2000) *Causality – Models, Reasoning and Inference*, Cambridge University Press
- 15 Halpern, J.Y. (2003) *Reasoning about Uncertainty*, MIT Press

Contributions to Monitor

We welcome recommendations of papers for review within *Monitor*, in the fields of combinatorial chemistry, pharmacogenomics, pharmacoproteomics, bioinformatics, new therapeutic targets, high throughput screening, new drug delivery technologies and other promising lines of research.

Details of recent papers or those *in press* should be directed to Dr Steve Carney, Editor, *Drug Discovery Today*, Elsevier Science London, 84 Theobald’s Road, London, UK WC1X 8RR. tel: +44 207 611 4132, fax: +44 207 611 4485, e-mail: DDT@drugdiscoverytoday.com