J Comput Aided Mol Des (2010) 24:957–960
DOI 10.1007/s10822-010-9388-7

PERSPECTIVE

Making priors a priority

Matthew Segall · Andrew Chadwick

Received: 28 September 2010 / Accepted: 1 October 2010 / Published online: 16 October 2010

Abstract When we build a predictive model of a drug property we rigorously assess its predictive accuracy, but we are rarely able to address the most important question, "How useful will the model be in making a decision in a practical context?" To answer this requires an understanding of the prior probability distribution ("the prior") and hence prevalence of negative outcomes due to the property being assessed. In this perspective, we illustrate the importance of the prior to assess the utility of a model in different contexts: to select or eliminate compounds, to prioritise compounds for further investigation using more expensive screens, or to combine models for different properties to select compounds with a balance of properties. In all three contexts, a better understanding of the prior probabilities of adverse events due to key factors will improve our ability to make good decisions in drug discovery, finding higher quality molecules more efficiently.

Keywords Prior · Availability bias · Decision tree · Multiparameter optimization

When developing a new computational model for predicting an activity or property of potential drug molecules we, quite rightly, focus on metrics for the accuracy of prediction; we usually report statistics such as the root-mean-square error, sensitivity, specificity or enrichment. To be rigorous, we report these statistics for independent test sets, so that the predictive power of our method can be assessed. If we are fortunate enough to work within a drug discovery project team, we also demonstrate the application of our model in a case study. However, we rarely ask the most important question of all; namely, how useful will this model be in making a decision in a practical context or, in other words, how predictive does the model need to be to make a difference to the decisions we make?

The value in applying a model comes from being able to select a limited number of molecules for progression, confident that those selected will meet the criterion for the predicted property. Or, conversely, through eliminating molecules from consideration, confident that molecules have not been incorrectly discarded, i.e. we have not lost opportunities to find a good drug. In order to address this question, we need an additional piece of information, the prior probability distribution of the property we are trying to assess.

A "prior" captures our understanding, or belief, of the likely outcomes of an event before the collection of new information (e.g. a measurement or prediction). The prior is a foundation of Bayesian probability theory, which defines how the probability of an outcome after an observation (the "posterior") can be inferred from the result of the observation, along with the prior for that outcome [1], as illustrated in an example below. In the case of a property of interest, the prior is the underlying probability distribution of the property at the time we wish to make a measurement or prediction, in the absence of new evidence.

In a recent paper [2], the authors discuss a number of psychological biases to good decision making in drug discovery. One of these, 'Availability Bias,' is the tendency to focus on recent events or results rather than looking at the sum of historical data and is sometimes described as "neglect of the prior". In this paper, drug discovery is

M. Segall (✉)
Optibrium Ltd, 7226 IQ Cambridge, Beach Drive, Cambridge CB25 9TL, UK
e-mail: matt.segall@optibrium.com

A. Chadwick
Tessella plc, Bretby Business Park, Ashby Road, Stanhope Bretby, Burton upon Trent, Staffs DE15 0YZ, UK

compared with evidence based medicine, so, in this spirit, let's consider a (fictitious) analogy from medicine to illustrate this issue:

In a given population, the prevalence of a fatal disease is 0.5%. Fortunately, there is a treatment for the disease although it is risky, carrying a 25% risk of mortality. There is a simple blood test for this disease and clinical trials have shown this to be 95% accurate (i.e. 95% sensitive and 95% specific). If a patient receives a positive test result, what is the appropriate course of action?

The first question is, "What proportion of the patients who test positive actually has the disease?" In this case the answer is approximately 9% (of 1,000 patients, on average $1{,}000 \times (0.005 \times 0.95 + 0.995 \times 0.05) = 54.5$ will test positive, but only 4.75 will have the disease). Therefore, even a positive test result is far from a death sentence. And, given the risk of the treatment, the rational decision is to do nothing. We would expect approximately 9% of those who test positive to die from the disease if we do nothing, but approximately 23% to die unnecessarily from the treatment!

The key point is that the utility of the test is critically dependent on the prevalence, i.e. the prior probability, of the disease. If this were to increase above 1.3%, the odds tip in favour of treatment. Looking at this in another way, with a prior probability of 0.5%, the test would have to be at least 98% accurate (assuming specificity and sensitivity were the same) to tip the odds in favour of treatment.

While this is a simplified example, similar situations present genuine challenges in medicine. For example, Welch and Black, in an analysis of overdiagnosis in cancer [3] note "Whereas early detection may well help some it undoubtedly hurts others. Often the decision about whether or not to pursue early cancer detection involves a delicate balance between benefits and harms… two out of three men with a raised PSA (prostate-specific antigen) will not have any cancer cells in their prostate biopsy". This type of analysis has, in practice, lead to restrictions on the use of diagnostic tests and procedures in some circumstances [4].

The analogous situation in drug discovery is a predictive model for, say, a rare toxicity. When it makes sense to 'kill' a compound or chemical series based on a prediction will depend on the interplay between the accuracy of the model and the prior probability of the risk factor. Unfortunately, at present we do not have good data on the priors for even the major causes of failure in pharmaceutical R&D. Therefore, we are in the difficult situation of not knowing when we may confidently base a decision on our predictions or when we may be misleading ourselves and at risk of falling foul of Availability Bias. Coarse data is available on causes of failure in development [5], but these group causes into broad categories, e.g. toxicity, PK, efficacy, and only covers the relatively small number of compounds that reach development and have therefore been heavily pre-screened. To have confidence, we require more detailed data on individual properties for larger numbers of compounds, including those in earlier discovery phases. Furthermore, the priors must be for unscreened compounds, at the point of application of the methods we are trying to evaluate; we would like to ascertain the probability of failure due to each cause in the absence of any further information or bias such as that revealed by screening or filtering. However, the prior is not for totally random compounds, as the collection of interest to be screened, would normally have been preselected in some way. Even a chemist's eye cast over a commercial compound catalogue is a form of reasoned preselection.

Of course, in drug discovery, computational methods are rarely used in isolation in the decision-making process. They are commonly used in conjunction with other methods, for example in vitro screens, that may offer greater accuracy in predicting the ultimate risk, although they are typically more expensive. Therefore, computational methods may be used to prioritise compounds for these more expensive screens before choosing compounds for further progression. However, here too, the prior probability of the risk factor, along with the reliabilities of the *in silico* and in vitro methods and their relative costs, determine the optimal strategy for combining these methods. An illustration of the interplay of these factors is illustrated in Fig. 1, which shows the relative values of five different strategies for combining *in silico* and in vitro screens for a given risk factor in one scenario. This example shows that a change in prior probability of a given adverse outcome from 30% to 40% would change the choice of optimal strategy from a sentinel approach [6] to one in which only the in vitro screen should be used. In the latter case, the inaccuracies of the *in silico* model, given the higher prevalence of the adverse event, mean that the cost of progressing compounds that are likely to yield expensive failures outweighs the saving from filtering out compounds that are likely to fail. An *in silico* model with greater specificity would be needed to add value to the screening cascade in this case. An interactive version of this simulation that explores the trade-off between accuracy and cost in the light of different prior probabilities in this two-screen cascade can be found at http://www.tessella.com/screening-strategy-explorer.

In addition, we are rarely concerned with a single risk factor in drug discovery; we need to balance many properties in a single molecule to find a successful drug including potency, selectivity, ADME, and toxicity [7]. Here, as illustrated in Fig. 2, methods for 'multi-parameter optimisation' can be used to prioritise compounds according to their likelihood of success, balancing the different criteria while taking into account the uncertainty in the sources of data and at the same time weighting by the

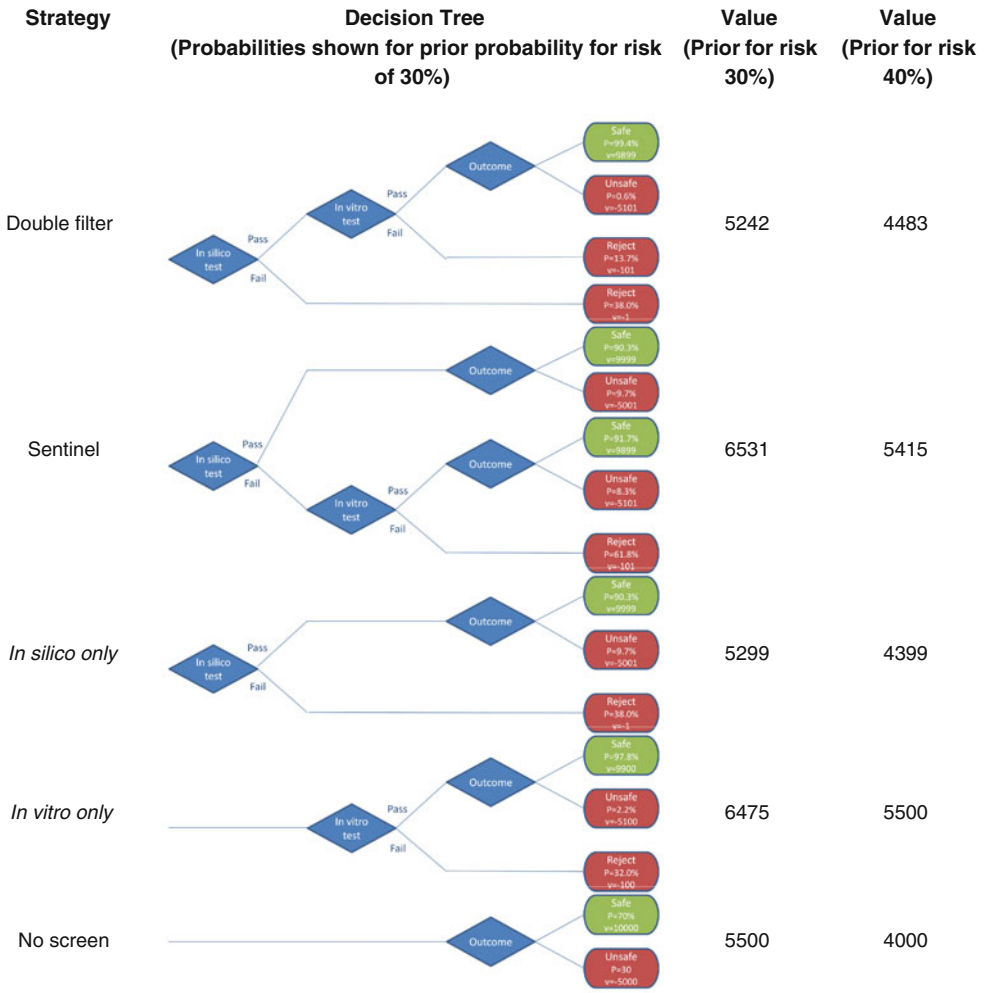

Fig. 1 Comparison of five different strategies for combining an *in silico* and in vitro screen for an adverse event. For each, the corresponding decision tree is shown and the expected value (arbitrary units) for two different prior probabilities of the adverse events, 30% and 40%. The parameters of the model were as follows: Accuracy of *in silico* screen 80% (specificity and sensitivity), Accuracy of in vitro screen 95% (specificity and sensitivity), cost per compound of *in silico* screen 1, cost per compound of in vitro screen 100, downstream cost to prove safety 5,000, net value of safe compound 10,000. Costs are in arbitrary units. Probabilities are shown in the leaves of the decision trees for a prior probability of 30% for illustration. The model assumes that the errors in the two methods are independent

relative risks associated with different factors [8]. Currently, assessment of these relative risks are based on subjective experience of experienced drug discovery scientists; however, here too, better priors would improve the quantitative assessment of the likelihood of success and allow compounds to be selected with greater confidence.

Conclusion

In order to get the most value from predictive models, they must be used to make effective decisions in drug discovery projects and hence help to find higher quality molecules more efficiently. A key to making better decisions using predictive models, and other screening methods, is to determine the prior probabilities for properties underlying the risks we are trying to eliminate in drug discovery. In principle, these data exist in the databases of pharmaceutical companies and analysis could yield the probability distributions that would inform improved decision-making across the industry. Many attempts have been made in the past to pool data between pharmaceutical companies, in particular to build better models of ADMET properties. These have foundered on the need to share proprietary compound structures or descriptors based on these structures, along with the data, in order to build models. To calculate priors, only the data itself is needed without associated structures. Perhaps this is a case where the industry can come together to provide the basis for better decision-making and hence improved efficiency and productivity for all?

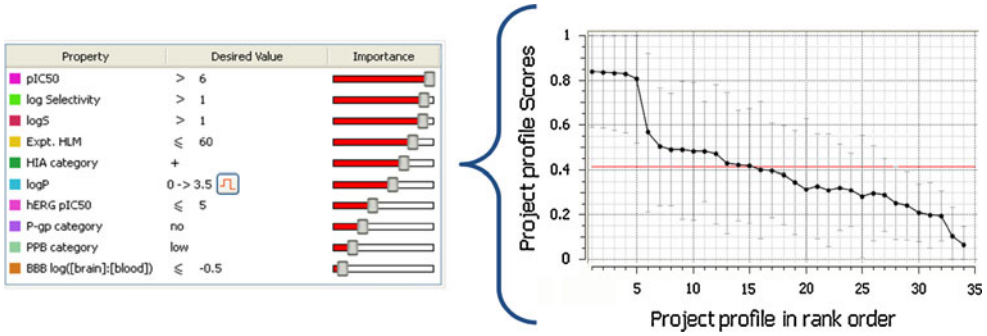

Fig. 2 Illustration of a probabilistic approach to multi-parameter optimisation, as implemented in the StarDrop™ software platform. A 'scoring profile' for a project defines the success criteria for predicted and experimental properties, indicating the ideal range for each property value ('Desired Value'), and relative importance (*slider bars*). The data for each compound, along with the associated experimental or statistical uncertainties, are assessed against this profile to calculate the likelihood of success (score) and overall uncertainty in the score. One approach to visualise this analysis is shown in the graph, in which compounds are plotted along the *x*-axis ranked from highest to lowest score. The score is plotted on the *y*-axis, with *error bars* indicating the overall uncertainty in the score. In this example, the top five compounds cannot be confidently distinguished; more data or criteria are required to choose between these. However, ~50% of compounds are significantly less likely to meet the project criteria than the top five and may be rejected with confidence

References

1. Jaynes ET (2003) Probability theory: the logic of science: principles and elementary applications vol 1. Cambridge University Press, Cambridge
2. Chadwick AT, Segall MD (2010) Overcoming psychological barriers to good discovery decisions. Drug Discovery Today 15:561–569
3. Welch HG, Black WC (2010) Overdiagnosis in cancer. J Natl Cancer Inst 102:605–613
4. COMARE (2007) 12th Report The impact of personally initiated X-ray computed tomography scanning for the health assessment of asymptomatic individuals. *Chairman:* Professor A. Elliott. COMARE Secretariat, Didcot. http://www.comare.org.uk/documents/COMARE12thReport.pdf. Accessed 7 June 2010
5. Di Masi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. J. Health Econ. 22:151–185
6. Performance assessment: has DEREK been improved? http://www.aapspharmaceutica.com/meetings/files/36/Kreatsoulas.ppt#302,14. Accessed 12 Oct 2010
7. Ekins S, Boulanger B, Swaan PW, Hupcey MA (2002) Towards a new age of virtual ADME/TOX and multidimensional drug discovery. J. Comp. Aided Mol Design 16:381–401
8. Segall MD, Champness E, Obrezanova O, Leeding C (2009) Beyond profiling: using ADMET models to guide decisions. Chem. Biodiv. 6:2144–2151