PERSPECTIVE

Benchmarking QSP Models Against Simple Models: A Path to Improved Comprehension and Predictive Performance

Andrew M. Stein[1]* and Michael Looby[2]

Quantitative Systems Pharmacology (QSP) models provide a means of integrating knowledge into a quantitative framework and, ideally, this integration leads to a better understanding of biology and better predictions of new experiments and clinical trials. In practice, these goals may be compromised by model complexity and uncertainty. To address these problems, we recommend that the predictive performance of QSP models be assessed through comparison with simpler models developed specifically for this purpose.

CPT Pharmacometrics Syst. Pharmacol. (2018) 7, 487-489; doi:10.1002/psp4.12311; published online on 22 Aug 2018.

Quantitative Systems Pharmacology (QSP) models are becoming more prevalent in the pharmaceutical industry.[1] These models provide a means of integrating knowledge into a quantitative framework and, ideally, this integration allows us to test our understanding of the system of interest and develop new hypotheses. An ultimate goal is to understand the system well enough to make accurate predictions of new experiments and clinical trials. In practice, these goals may be compromised by inherent model complexity and uncertainty. The complexity of models (including QSP models) often leads to overfitting of noise in the data rather than capturing true knowledge of the system, thereby compromising predictive performance. In many fields, it has been demonstrated that simple heuristics outperform complex models in terms of predictive performance[2] and this suggests the complex models may not capture important system features accurately or precisely enough due to overfitting. For these reasons, we recommend that a QSP modeling plans should include development of simpler models to assess the predictive performance of the QSP model in specific contexts. In this perspective, we use examples from pharma and other fields to support this recommendation.

Simple models for comprehension

Often, a complex, mechanistic model can be simplified by focusing on steady state, lumping compartments, and using approximations. Simplification helps the model developer to build understanding and intuition of the QSP model, and also can lead to clearer presentations of the model and results. Simplification allows one to focus on the model elements that drive the predictions of interest, while also providing insights into which aspects of the complex mechanistic model provide explanatory power. Three examples of the benefit of simplifying complex models are provided below, with the obvious caveat that not all questions are answerable with a simplified model.

Physiologically based pharmacokinetic models

Pharmacokinetic modeling is the classic example of using a simple model to describe a complex system. Physiologically based pharmacokinetic models have many applications, including lead optimization and prediction of drug-drug interactions. However, for biologics, there is not yet consensus on the model structure and model parameters. Fronton *et al.*[3] have demonstrated how parameter lumping can allow one to relate all parameters in the simple pharmacokinetic model to the more complex physiologically based pharmacokinetic model, such that consensus on a more detailed parametric model is not necessary for addressing questions about tissue distribution at large, target-saturating doses.

Target engagement and the Hill equation

The Hill equation was one of the first quantitative, parametric models that related the response of a system to the concentration of a pharmacological agonist.[4] Subsequent receptor models generally incorporated the Hill equation, typically in more complex formulae. Further research revealed that the parameters of the Hill equation (or its simple derivatives) only have a precise physicochemical meaning when applied to ligand binding reactions (especially when the Hill coefficient is unity). However, in contrast with the advanced models developed since, the Hill equation requires fewer assumptions and less *a priori* knowledge about the mechanism of action to provide useful results. The simplicity, flexibility, and reliability of the Hill equation make it an essential starting point for exposure-response analysis in physiological and pharmacological investigations and a starting point for the development of new pharmacological models. Insights from the Hill equation have also been extended *in vivo*, to include receptor turnover.[5]

Human immunodeficiency virus replication

Human immunodeficiency virus (HIV) typically takes about 10 years to advance from initial infection to full-blown AIDS; this suggested that the rate of HIV replication was very

[1]Novartis Institute for Biomedical Research, Cambridge, Massachusetts, USA; [2]Novartis Pharma AG, Basel, Switzerland. *Correspondence: Andrew M. Stein (andrew.stein@novartis.com)

Received 23 March 2018; accepted 3 May 2018; published online on 22 Aug 2018. doi:10.1002/psp4.12311

slow. During clinical trials of ritonavir, a drug that stopped HIV replication, it was observed that HIV RNA levels dropped rapidly, with a half-life of about 2 days. Mathematical models of the HIV infection of CD4+ cells were available at the time of analysis, but a simple expression of mass balance was sufficient to show that the rapid drop in HIV RNA levels after therapy meant that, before therapy, the replication rate was also fast, and, therefore, combination therapy would be needed to prevent the rapid emergence of resistance to a monotherapy.[6]

Simple models for prediction

"The test of science is its ability to predict."
- Richard Feynman

Although a QSP model can have value beyond prediction (e.g., evaluating whether a mechanistic hypothesis is consistent with data or charactering the most sensitive components of a system), the strongest proof that a model is accurate is to show prospectively that it can make more accurate predictions than other methods. Demonstration of predictive accuracy provides convincing evidence to a broad audience. For example, after Newton's laws of motion were used by Halley to predict the appearance of a comet, the power of these laws could then be demonstrated to people without a background in physics or mathematics.

There are many limitations of QSP that can make prediction challenging:

1. Limitations in biological understanding due to: (a) positive and negative feedback loops (e.g., cell signaling pathways, immune response)[7]; (b) system evolution over time (e.g., immune response to antigen or resistance acquisition in oncology); (c) large heterogeneity, unexplained variability, and/or sensitivity to initial conditions (e.g., presence of small resistant clone in a tumor, presence of T cell epitope that recognizes a particular antigen); and (d) significant gaps in the understanding of the underlying biology and physiology, in particular "unknown, unknowns" (i.e., properties of the system that we do not understand or even know exist, and, therefore, cannot be included in the model).[7]
2. Limitations in measurement accuracy due to: (a) poor markers that do not capture system behavior in a consistent or precise manner; (b) cell line misidentification; (c) irreproducible experiments; (d) unvalidated assays; and (e) "file-drawer" effect from researchers publishing positive results more frequently than negative results, which leads to a bias in the available data.
3. Limitations in model accuracy due to: (a) practical unidentifiability in the parameters; (b) significant uncertainty in model parameter; and (c) structural model error.

Many of the above limitations are often simultaneously at play, which greatly increases the risk of overfitting, making predictions inaccurate. Given the significant resource necessary to develop complex models, it is essential to apply a strategy to assess overfitting at the model development stage. One approach to assess overfitting is by comparing the predictive performance of complex mechanistic models to simple, context-specific models. Three examples are given below.

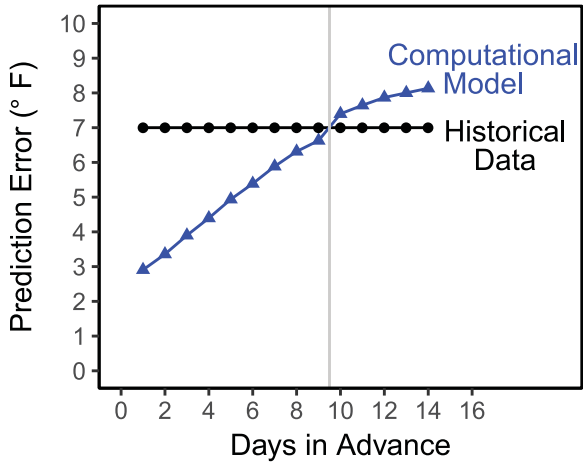

Figure 1 A comparison of two types of weather predictions. One uses historical data alone and the other uses a computational model. The computational model works best over short time scales and the simple model outperforms the computational model at making predictions more than 10 days into the future.

Daily temperature forecasts

We start with an example outside QSP in which the data is rich and allows for a clear comparison between a simple and complex model. To predict the daily temperature, a simple model uses the historical records of the average daily temperature in years past to make a prediction for the day of interest, whereas the complex model makes the temperature prediction using a computational model of the atmosphere. As shown in **Figure 1**, the computational model outperforms the historical record only when the prediction is <10 days into the future.[8] When extrapolating further, the simpler approach using only historical data is more accurate. With QSP models, the goal is often to predict how a drug will work in humans based on an integration of preclinical and clinical data. A useful question to consider is whether enough is known about the system for making accurate predictions of clinical outcomes with a QSP model, or whether making such predictions is just as difficult as predicting the weather more than 10 days in advance. In some settings, simpler methods may make more accurate predictions.

Cardiotoxicity

To predict Torsades de Pointes from *in vitro* ion channel data,[9] the complex approach used biophysical models describing the change in ion-channel conductance over time within a cardiac cell using hundreds of parameters and dozens of differential equations. The simple approach used a single parameter, which was calculated by looking at the *in vitro* data of the clinically relevant drug concentration, adding the percent block of each repolarizing ion channel, and then subtracting the sum of the percent block of each polarizing ion channel. It was found that the simple model performed just as well and sometimes better than

the biophysical models. The conclusion is that although the biophysical models may have many applications, for predicting Torsades de Pointes, the key factors of interest are the polarization and depolarization of the relevant ion channels, and these are adequately captured with the simple modeling approach.

Oncology combinations

There is currently much effort devoted to developing QSP models to predict how immunotherapy agents should be combined with each other or with other agents to maximize benefit to patients. To our knowledge, there has not been any effort yet to compare QSP model predictions to simpler approaches. However, a simple approach is available: starting from monotherapy response data, assume each drug acts independently or with a small degree of correlation, and then use a probabilistic model to predict the outcome of the combination. It has been shown that this simple approach can predict the response of a significant number of combination studies.[10] Going forward, it is recommended that QSP efforts for predicting combination response be compared to this simple heuristic. If QSP oncology models make better efficacy predictions over a large number of examples, then the value of these models can be clearly demonstrated to a broad audience. However, if the simpler approaches work better, this means that there are important underlying processes that are not yet characterized by the model or there is a problem of overfitting.

CONCLUSIONS

We propose when developing a QSP model for the purpose of making predictions for a large number of drugs (e.g., cardiotoxicity, hepatotoxicity, and oncology drug combinations), benchmarking this model against simple, context-specific heuristics is necessary to assess potential overfitting and the resulting degradation in predictive performance.

Acknowledgments. The authors would like to thank Leon Aarons, Peter Bonate, Dean Bottino, Lulu Chu, Jaeyeon Kim, Alison Margolskee, Hitesh Mistry, Andreas Raue, and Jean-Louis Steimer for helpful discussions.

Source of Funding. No funding was received for this work.

Conflict of Interest. A.S. and M.L. both work for Novartis.

1. Danhof, M. Systems pharmacology – towards the modeling of network interactions. *Eur. J. Pharm. Sci.* **94**, 4–14 (2016).
2. Haldane, A.G. The dog and the frisbee. Speech given at the Federal Reserve Bank of Kansas City's 36th Economic Policy Symposium, "The Chang Policy Landscape," Jackson Hole, Wyoming, USA. (31 August 2012).
3. Fronton, L., Pilari, S. & Huisinga, W. Monoclonal antibody disposition: a simplified PBPK model and its implications for the derivation and interpretation of classical compartment models. *J. Pharmacokinet. Pharmacodyn.* **41**, 87–107 (2014).
4. Gesztelyi, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B. & Tosaki, A. The Hill equation and the origin of quantitative pharmacology. *Arch. Hist. Exact Sci.* **66**, 427–438 (2012).
5. Stein, A.M. & Ramakrishna. R. AFIR: a dimensionless potency metric for characterizing the activity of monoclonal antibodies. *Clin. Pharmacometrics Syst. Pharmacol.* **6**, 258–266 (2017).
6. Perelson, A.S. Modelling viral and immune system dynamics. *Nat. Rev. Immunol.* **2**, 28–36 (2002).
7. Orrell, D. *The Future of Everything: The Science of Prediction* (Thunder's Mouth Press, New York, NY, 2008).
8. Silver, N. *The Signal and The Noise: Why So Many Predictions Fail—But Some Don't* 1st ed (Penguin Press, New York, NY, 2012).
9. Mistry, H.B. Complex versus simple models: ion-channel cardiac toxicity prediction. *PeerJ.* **6**, e4352 (2018).
10. Palmer, A.C. & Sorger, P.K. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell.* **171**, 1678–1691 (2017).