COMMENTARY
How Do We “Validate” a QSP Model?

Daniel C. Kirouac

This is a question that arises recurrently, both in the process of publishing Quantitative Systems Pharmacology (QSP) research – and more critically for those making drug development decisions based on model simulations. Although previous commentaries have addressed this issue, there is still no agreed-upon answer. Using the associated article by Zhu et al. as an example, I will describe how Virtual Populations can be used to assess confidence in qualitative predictions from QSP models.

QSP models are more than multicompartment pharmacokinetic/pharmacodynamic

Pharmacometrics, in particular nonlinear mixed effect (NLME)-based population pharmacokinetic/pharmacodynamic modeling, has established methods for model validation, such as goodness-of-fit and visual predictive check plots. However, these do not necessarily translate directly to QSP models due to differences in both the intended purpose and the data used in model development. Although dose-exposure-response projections are a common motive for developing QSP models, the scope can be much broader. This may encompass novel target and biomarker identification, predicting mechanisms of drug resistance, and unravelling the biological processes underlying complex behavior. These constitute qualitative predictions, and, thus, are not necessarily compatible with methods developed for longitudinal (pharmacokinetic) analyses. Biological plausibility often takes precedence over model parsimony in QSP models. As such, the models are often larger, and contain multiple nonlinearities and nonidentifiable parameters. Precision of parameter estimates is, therefore, a much less important concern, and is often unimportant. As a result, model selection criteria (such as Akaike or Bayesian information) are not necessarily relevant.

The data used to constrain QSP models differs from pharmacometrics as it is often assembled from disparate sources rather than a single curated dataset. These may span multiple biological scales, covering in vitro, in vivo, and clinical assays, and may be qualitative (e.g., increase vs. decrease, a rank ordering of effects). As such, k-fold cross-validation methods are often inappropriate.

Despite these differences, prediction remains the fundamental goal of any computational model. Can we predict phenomena (e.g., the efficacy of a novel dosing regimen) on a computer, bypassing the need to conduct costly and time-consuming experiments? How then do we validate QSP models to gain confidence in the predictions? It is actually less challenging than one might think if we focus on the fundamental concepts rather than methodology.

Qualitative predictions from QSP models

Predicting qualitative relationships is often a primary objective of QSP models, and the work presented by Zhu et al. in this issue, provides a good example. The authors developed a relatively comprehensive model of intracellular signaling pathways regulating pancreatic cancer cell proliferation and survival in the context of two clinically relevant drugs: the nucleoside analogue gemcitabine, a standard-of-care chemotherapy, and birinapant, an antagonist of IAP-family proteins that induces sensitivity to apoptotic stimuli (such as gemcitabine). The system of differential equations was parameterized using dynamic protein measurements from a cell line (PANC-1) following drug treatment, and the protein dynamics linked to cell cycle progression, apoptosis, and cellular kinetics. Model simulations subsequently “predicted” results from the published literature that were not used in model training or explicitly encoded in the structure.

Two predictions related to drug combinations are particularly relevant. The researchers identified a drug-scheduling effect, wherein sequential treatment of gemcitabine followed by birinapant was more efficacious than simultaneous treatment. As well, the natural nuclear factor-kappa β inhibitor curcumin, an agent that was not included in model training, was predicted to have less-than-additive activity in combination with gemcitabine. Both constitute nonintuitive and potentially clinically actionable results. Although these effects are apparent from examining the simulated cell growth curves, how could one further quantify the statistical robustness of these findings?

Using virtual populations to quantify qualitative results

At heart, this involves accounting for both uncertainty and variability. In NLME-based population models, this is accomplished using fixed (population-based) and random (subject-based) parameters. The concept of virtual populations (VPs), although ill-defined, is gaining traction in fulfilling this role in QSP. Although no universally agreed-upon terminology exits, “Virtual Subject” generally refers to a single model parameterization, “Virtual Cohort” a family of model parameter sets, and “Virtual Population” a family of parameter sets weighted to match a clinical or other response distribution. Practically, this comprises a parameter matrix (parameters × subjects) with relative weights tied to the prevalence of each subject in the population. These can be generated by running multiple iterations of a stochastic parameter optimization algorithm, by repeated sampling around some distribution of a single parameter set, or some mixture of the two. Due to the
nonlinear complexity of QSP models, it is common for response predictions to be tightly constrained while the underlying parameters span a wide range.4

The value of the VP approach lies in generating distributions of predictions. This is useful not just in the conventional means of assigning confidence intervals to simulated trajectories, but in qualitative analyses as well. For example, one could ask in what proportion of VP simulations is a specific target identified as critical, or a given type of input-output relationship observed. A null hypothesis for comparison can then be created by running model simulations drawn from random parameter sets, or with random perturbations (i.e., drug treatments). Discrete statistical methods can then be used to compare the two distributions and evaluate the likelihood of such qualitative predictions (drug targets, combination effects, mechanisms of resistance, etc.) arising by chance alone.

Although it precedes the term VP, this concept has a long history in computational systems biology. Using cancer cell signaling as an example, ordinary differential equation-based models are often characterized by a family of parameter estimates, which can then be used to generate a distribution of model simulations.7,8 Although less common, related approaches have been applied to logic-based qualitative models. By examining the distribution of parameter estimates from multiple optimization runs, one can weigh the importance of intracellular connections, discern between alternate putative mechanisms, and predict the activity of novel drug targets or drug combinations.9,10

Zhu et al.2 could have applied a VP approach to quantify the robustness of their predictions. For example, they could have asked in what proportion of simulations does one observe the gemcitabine-birinapant schedule effect, or the gemcitabine-curcumin less-than-additive combination activity. This could be compared to a null hypothesis, simulations drawn from random parameter sets, or from “drugging” random pairs of proteins, to assess significance.

The challenge is that no established methods exist for doing so, and there are few off-the-shelf algorithms or ready-to-implement code. Performing this kind of analysis is thus computationally intense, and involves subjective decisions about how to implement and interpret the results. The lack of consistency in QSP model validation criteria is therefore understandable. However, quantifying the statistical significance of results is critical to good science, and should be an expected component of QSP.

Charting a path forward
One of the most frequent criticisms of QSP models is that they contain so many free parameters that one could fit anything with them. Anyone who has struggled in vain to force a model to fit unyielding data will recognize this as a falsehood. Nevertheless, systematic approaches for exploring parameter space and quantifying model accuracy will reduce these kinds of criticisms and help advance the field toward broader acceptance.

No standard operating procedures exist for QSP, which is what makes the work both challenging and exciting. Developing such models is, thus, akin to exploring uncharted territory, and we need some navigational tools to keep from getting lost.

Acknowledgments. The author thanks Joshua Apgar, Lore Gruenbaum (Applied BioMath), and Richard Berman (VerbFactory) for helpful feedback.

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

Conflict of Interest. D.C.K. is an employee and shareholder of Applied BioMath LLC.

1. Agoram, B. Evaluating systems pharmacology models is different from evaluating standard pharmacokinetic-pharmacodynamic models. CPT Pharmacometrics Syst. Pharmacol. 3, e101 (2014).
2. Zhu, X., Shen, X., Qu, J. & Justko, R.M. Multi-scale network model supported by proteomics for analysis of combined gemcitabine and birinapant effects in pancreatic cancer cells. CPT Pharmacometrics Syst. Pharmacol. (2018).
3. Friedrich, C.M. A model qualification method for mechanistic physiological QSP models to support model-informed drug development. CPT Pharmacometrics Syst. Pharmacol. 5, 43–53 (2016).
4. Allen, R.J., Rieger, T.R. & Musante, C.J. Efficient generation and selection of virtual populations in quantitative systems pharmacology models. CPT Pharmacometrics Syst. Pharmacol. 5, 140–146 (2016).
5. Gadkar, K., Kirouac, D.C., Mager, D.E., van der Graaf, P.H. & Ramanujan, S. A six-stage workflow for robust application of systems pharmacology. CPT Pharmacometrics Syst. Pharmacol. 5, 235–249 (2016).
6. Kirouac, D.C. et al. Clinical responses to ERK inhibition in BRAFV600E-mutant colorectal cancer predicted using a computational model. NPJ Syst. Biol. Appl. 3, 14 (2017).
7. Chen, W.W. et al. Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol. Syst. Biol. 5, 239 (2009).
8. Gaudet, S., Spencer, S.L., Chen, W.W. & Sorger, P.K. Exploring the contextual sensitivity of factors that determine cell-to-cell variability in receptor-mediated apoptosis. PLoS Comput. Biol. 8, e1002482 (2012).
9. Morris, M.K., Clarke, D.C., Osimiri, L.C. & Lafuentenburger, D.A. Systematic analysis of quantitative logic model ensembles predicts drug combination effects on cell signaling networks. CPT Pharmacometrics Syst. Pharmacol. 5, 544–553 (2016).
10. Eduati, F. et al. Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. Cancer Res. 77, 3364–3375 (2017).

© 2018 The Authors CPT: Pharmacometrics & Systems Pharmacology published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.