COMMENTARY

Negative Modeling Results: A Dime a Dozen or a Stepping Stone to Scientific Discovery?

B Hendriks1

CPT: Pharmacometrics & Systems Pharmacology (2013) 2, e48; doi:10.1038/psp.2013.28; advance online publication 12 June 2013

Systems pharmacology models, in general, tend to span multiple timescales bridging detailed mechanism with higher level responses or functional outputs. These features serve to put systems pharmacology models in a separate class that brings with it specific challenges, particularly in evaluating such work. When models are constructed on well-understood mechanisms but fail to match experimental data, in what cases should "negative modeling results" be considered scientific findings?

I will start with the following premise: systems pharmacology is an approach to facilitate understanding pharmacology (as opposed to an end in itself). Therefore, there must be an interplay between models and experimental results. This can be observed to occur on two timescales (i) inter-publication, where a group publishes experimental results which are then retrospectively analyzed by a computational group or (ii) intra-publication, wherein the same group performs both activities and the outcome of one activity directly feeds into and motivates the other. Systems pharmacology models, in general, tend to span multiple timescales bridging detailed mechanism with higher level responses or functional outputs. Parameter values tend to be justified based on previous work or trained on a small scale data rather identified as part of a large-scale optimization. These features serve to put systems pharmacology models in a separate class that brings with it specific challenges.

As an example, the study by Darwich et al.1 entitled “Evaluation of an in silico PBPK post bariatric surgery model through simulating oral drug bioavailability of atorvastatin and cyclosporine“, presented in this issue, takes a well-established systems pharmacology model and tests its simulations against available clinical data for various types of bariatric surgery. I believe it is fair to say that the overall model performance is mixed, judged by its ability to predict the clinical data, depending on the drug and type of surgery. This has prompted discussion on the value, relevance, and interest of negative modeling results and stimulates one to consider a larger question as well: How should one evaluate systems pharmacology efforts and when are they worthy of publication?

SIGNIFICANCE

In the systems pharmacology domain, the significance of most work is driven by clinical need. This is usually a clinical observation that impacts patient care wherein additional information would change clinical decision making and/or outcomes. Significance is also often derived from enabling decision making in the drug development space. Pure academic pursuit may also motivate some work wherein systems level thinking yields new insight into a basic understanding of pharmacological and/or biological mechanism or in the form of novel methods. Finally, significance may be derived from the integrative nature of bringing together a vast body of knowledge into a coherent and quantitative framework to determine if the pieces “add up.”

For Darwich et al., there is a good case to be made that there is a great clinical significance to understanding how various weight loss surgeries may affect drug exposure. With a paucity of clinical data, and the absence of large patient populations having undergone surgery, simulation using systems pharmacology models is a very promising manner by which to begin filling the knowledge gap.

MODEL VALIDATION: CREDIBILITY VS. EVALUATION

In developing a model, its credibility is based on assumptions made in its structure and the extent to which those are well accepted in the field. Second, the degree of detail, or modeling abstraction, will also determine the ease with which it is accepted by others. Credibility is also based on how believable the simulation results are to specific observations for key predictions. For better or worse, the aspect of credibility is largely a subjective one.

In communicating computational models, it is a good scientific practice to present the grounds by which one measures the model to be a valid representation of the system modeled. In the absence of data, one might have the “luxury” to build on reasonable assumptions and explore the effects, but impact is limited by the lack of appropriate data and metrics to judge quality. With limited data, the model development process usually entails evaluating multiple competing mechanisms, hypotheses, and/or representations and usually only what worked best makes it to publication. In these cases, one can merely assert that the model, with its assumptions, is “consistent with experimental observations.” One cannot prove actual mechanism although one can effectively discredit proposed mechanisms as inconsistent with data. A model that captures data well can be used to generate subsequent testable hypotheses using the model. With additional

1Merrimack Pharmaceuticals, Cambridge, Massachusetts, USA. Correspondence: B Hendriks (bhendriks@merrimackpharma.com)
Received 20 March 2013; accepted 3 April 2013; advance online publication 12 June 2013. doi:10.1038/psp.2013.28
data, these models are generally judged to be more credible. Some prefer parsimonious models to reduce complexity and aid in the ability to determine (or in some cases mathematically identify) parameter values, but this may come at the expense of incorporating known mechanisms. If you push any model hard enough—i.e., challenge it with enough data and varied conditions—you can find places where it breaks. This is necessarily true as these models are all simplified representations. There is then a fine line between model validation testing, repurposing, and ultimately failure. When an established model fails at performing a new task—is this failure of the model or a finding? Under what scenarios is such a finding worthy of wider dissemination?

Again, I will refer to the study by Darwich et al. as an example. Their work used a previously published model, indicating that the structure and assumptions have been reviewed and accepted following peer review. This suggests that the model is credible. For evaluation of the model, the authors relied largely on qualitative comparisons between model simulation and clinical data. In a few cases, they changed parameters from their nominal values to explore their ability to better describe the data, once in a directed manner and the other as part of a systematic sensitivity analysis. What measures should be used to judge the quality of the model predictions against the data? In cases where the model prediction differed from the clinical data should this be viewed as a short-coming of the model’s credibility or is it a scientific finding?

THE VALUE OF NEGATIVE (MODELING) RESULTS

Negative results are generally the stepping stones to scientific discovery. The value of the model development process is frequently enlightening but rarely presented. There are some scenarios where publishing negative findings may be instructive. For this to be the case, it is absolutely imperative to have a model built on credible assumptions and implemented with technical accuracy to distinguish itself from a null result (a failure of technical implementation) where nothing is learned. This is increasingly difficult to evaluate (for positive and negative results), as models grow in scale and complexity.

In cases where negative results can effectively discredit commonly held dogma, negative results can be useful and of widespread interest. More specifically, presenting plausible models that are not consistent with experimental or clinical findings can be useful. In practice, there is a biphasic relationship between model complexity and the ability to effectively communicate results. If the model is too small and simple, it will likely be discredited for being over simplified. It is expected that as it increases in complexity, it will gain explanatory power, but at some point, it may become discredited for being too difficult to evaluate and communicate the model structure. It is anticipated that there is a sweet spot of complexity with regard to presenting both negative and positive modeling results.

Practically speaking, however, dogma can usually not be unseated without presenting a new explanation to take its place. In this manner, negative results are typically complemented by positive findings that extract learning from negative results. This comes first in the form of hypotheses generated and followed by computational and/or experimental testing. Therefore, if one seeks to publish negative results, one should be able to clearly articulate what has been learned from the negative results, why the negative results are of significant interest to the community, and what hypotheses they suggest. Did Darwich et al. sufficiently succeed in this regard?

EXPECTATIONS OF SYSTEMS PHARMACOLOGY

If the main purposes of publishing is to share knowledge and enable others to build on one’s work, what should be the metric(s) by which to judge high-quality work? Certainly, work needs to be technically strong (in a computational and/or experimental sense) with reasonable and justifiable hypotheses. How good does a model need to be able to recapitulate data to be considered worthy of publication? Should high significance or unmet need affect the standards by which work should be judged? Or is it more about interpretation and presentation of those results? What is the right balance between communicating the learning event (the negative result) or the subsequent insight (the arising hypothesis and subsequent testing)? As an emerging field, these are questions that will need to be answered and may hold the key to generating and disseminating high-impact work.

Acknowledgments. The author acknowledges helpful discussions from Piet van der Graaf and Stephen Duffull.

Conflict of Interest. B.S.H. is employed by and holds stock and/or stock options in Merrimack Pharmaceuticals.

1. Darwich, A.S. et al. Evaluation of an in silico PBPK post bariatric surgery model through simulating oral drug bioavailability of atorvastatin and cyclosporine. CPT: Pharmacomet. Syst. Pharmacol. 2, e47 (2013).