Lessons Learned from Model Development that Seemed to Go Wrong

Nina Isoherranen, PhD

SCHOOL OF PHARMACY
UNIVERSITY OF WASHINGTON
References and Acknowledgments

Weize Huang, PharmD, Graduate student

Funding: NIH NIDA P01 DA032507, Warren G Magnuson Scholarship (WH) and William E Bradley Fellowship (WH) from UW

1. Huang and Isoherranen, Development of a Dynamic Physiologically Based Mechanistic Kidney Model to Predict Renal Clearance. CPT:PSP 2018,7:593-602
2. Huang and Isoherranen, Sampling Site Has a Critical Impact on PBPK Modeling. JPET 2019, doi: 10.1124
The concept of fit-for-purpose modeling

To be useful, PBPK modeling should provide data that help with decision making at the appropriate stage of drug development

OR

Drive hypothesis generation of mechanisms or processes that can be tested in experiments- A model should help with understanding of the system

“Is the model used simply the “best available” at the present time, or is it truly adequate for the specific purpose of interest? How would adequacy for purpose be assessed, and what would it look like”? (Thompson and Smith, 2019)
Simulations that seem to go horribly wrong

Several hypotheses what could be “wrong”
1. Something is not right with the solver/how the simulation is running
2. The structural model is not right
3. Drug model is somehow flawed
By “optimizing” the $K_p$ values we can get a curve that overall agrees with observations.

Increasing liver methamphetamine $K_p$ value from predicted 3.3 to 13.3 starts to resolve the shape issue of the amphetamine concentration curve (CL fixed).
The degrees of freedom and level of uncertainty in $K_p$ values

- $K_p$ values can be predicted (commonly done with Rodgers & Rowland method) but is this good enough?
- “distinction between simulated variables and their real-world counterparts can become unclear” (Thompson and Smith, 2019)

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Predicted using R, L &amp; R eq with rat-human hybrid</th>
<th>Predicted using commercial software</th>
<th>Observed from PET data (Volkow et al 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>1.85</td>
<td>1.82</td>
<td>9.67</td>
</tr>
<tr>
<td>Heart</td>
<td>2.28</td>
<td>1.44</td>
<td>5.21</td>
</tr>
<tr>
<td>Kidney</td>
<td>2.13</td>
<td>1.55</td>
<td>14.55</td>
</tr>
<tr>
<td>Liver</td>
<td>3.02</td>
<td>1.65</td>
<td>24.96</td>
</tr>
<tr>
<td>Lung</td>
<td>1.58</td>
<td>1.51</td>
<td>6.94</td>
</tr>
</tbody>
</table>
Using PET-scan \( K_p \) values metabolite curve looks acceptable

- Is this model good enough? Fit-for-purpose?
- Does it replicate observed data for the wrong reasons?
- Conclusion that the drug model was the cause for the unacceptable simulation?
Structural model may still need reassessment

- PBPK models typically sample from merged venous compartment - analogous to right atrium
- Observed PK is usually from an arm vein
- Does the difference in sampling sites matter?

![Diagram of PBPK model](image)

<table>
<thead>
<tr>
<th>Arterial and Venous Conc. (µg/L)</th>
<th>Ketamine 7 mg IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (hr)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ AAFe_{central} = 1.26 \]

\[ AAFe_{peripheral} = 1.20 \]
The model structure and/or discrepant sampling site may be the reason for model rejection

- The metabolite profile is different from different sampling sites
- Optimizing a model to discrepant sampling site may lead to wrong parameters
- For different compounds/scenarios different parts of the model may be sensitive
- Does this impact extrapolation to unstudied scenarios
Momentsous sprint at the 2156 Olympics?

Women sprinters are closing the gap on men and may one day overtake them.

“a far more interesting race should occur in about 2636, when times of less than zero seconds will be recorded.” Rice K, Nature, 2004

“Since all models are wrong the scientist must be alert to what is importantly wrong” “we make tentative assumptions about the real world which we know are false but which we believe may be useful nonetheless” Box G, JASA, 1976
Observed arterial and venous concentrations are different. Arteriovenous difference can be captured by PBPK. Each site independently captures the corresponding plasma concentrations-time curves.
Fit for purpose re-examined: Buccal fentanyl model optimization

Arm vein sampling

Central PBPK sampling

Fitting central model into observed arm vein data

$\begin{align*}
    k_{a,buccal} &= 1 \text{ hr}^{-1} \\
    k_{a,gut} &= 3 \text{ hr}^{-1} \\
    k_{a,buccal} &= 0.35 \text{ hr}^{-1} \\
    k_{a,gut} &= 0.7 \text{ hr}^{-1}
\end{align*}$

> Is there a possibility that “best available” model, while providing accurate recapitulation of observed data (for wrong reasons), results in inaccurate extrapolation
Extrapolation built on wrong optimization could mislead decision making

> Seemingly unimportant (minor) model optimizations can have a major impact on predictions of unstudied scenarios

“Lorenz’s butterfly effect”

> "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful." Box G, 1987
The number of parameters and degrees of freedom - structural and drug models

- PBPK models have multiple compartments, usually at least 14
- Each compartment is associated with several parameters, for example: blood flow, organ volume, $K_p$, $CL_{distribution}$, $CL_{int}$
- At least 70 degrees of freedom (10$^{70}$ simulations needed for global sensitivity analysis and 700 for local sensitivity analysis if testing 10 values per parameter): **identification of sensitive parameters and strategizing sensitivity analyses for particular model can be challenging**
- Statistical dilemma; how many datapoints would be needed to differentiate one structural model from another?
- In many cases, the optimization of multiple parameters is interdependent
Models to predict renal clearance have several interdependent parameters

Methamphetamine and amphetamine have pH-dependent passive tubular reabsorption and active secretion transport.

- From observed data tubular secretion and passive reabsorption are interdependent processes.

Acidic urine pH 4.9-5.3

Uncontrolled urine pH

Alkaline urine pH 7.8-8.2
Development of Mechanistic Kidney Model to Gain Confidence on Passive Reabsorption

- Model performance verified with number of drugs that do not have active secretion
Application of Verified Mechanistic Kidney Model to Predict Urine pH effect on Amphetamine disposition

While different physiological scenarios can be simulated, their probability to be seen in a sample cannot be estimated without the knowledge of parameter distribution.
Conclusions

> K_p values should always be experimentally determined and V_{ss} confirmed with iv dosing data

> The sampling site in PBPK simulations should be matched to the sampling site in experimental studies or vice versa

> The level of complexity needed in PBPK models can vary. It is critical to determine the necessary complexity and aim for parsimony. This will allow identification of components in the model that are importantly wrong (ie make model not useful)

> Statistical methods that account for increased degrees of freedom in “better fits” are needed to differentiate models

> Extrapolation to unstudied scenarios makes an implicit assumption that the same mathematical relationships apply from interpolation through extrapolation