2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

CRISP: A Modeling Framework for Identifiability and Interpretation in PBPK-Based Rpt Dosimetry

Christian N. K. Anderson[1]; Jackson Pond[1]; James Fowler[2,3]; Maziar Sabouri[2,3]; Taylor J McColl[2,3]; Tahir Yusufaly[4]; Arman Rahmim[2,3]; Carlos Uribe[2,3]; Mark K. Transtrum[1]
[1]Cross Stream Bioanalytics, [2]BC Cancer Research Institute, [3]University of BC, [4]Radiology Dept, Johns Hopkins University

ABSTRACT

Purpose: Physiologically based pharmacokinetic (PBPK) models used in radiopharmaceutical therapy (RPT) and theranostic digital twins are rapidly increasing in size and mechanistic detail, often containing tens to hundreds of parameters. While such models fit imaging and dosimetric data well, their high dimensionality pose fundamental challenges for parameter identifiability, uncertainty quantification, and mechanistic interpretation. This work introduces Contextualized Reduction for Identifiability and Scientific Precision (CRISP), an analysis framework developed in quantitative systems pharmacology (QSP), as a transferable approach for addressing these challenges in PBPK-based theranostics.

Methods: CRISP was developed to analyze "sloppy" models, i.e., models whose predictions are controlled by a few effective parameter combinations while remaining insensitive to many others. The framework uses Fisher Information analysis, information geometry, and systematic model reduction using the Manifold Boundary Approximation Method (MBAM). CRISP has been successfully applied in QSP, where sloppiness similarly bottlenecks traditional analysis. Here, PBPK-based theranostics (TAC) is fitting is similarly examined. We consider a 90-Y/111-IN PBPK compartment model with 179 parameters (65 independent), representative of emerging theranostic digital twins, to illustrate properties of sloppiness under realistic data constraints.

Results: Large PBPK models for RPTs exhibit wide separation of parameter sensitivities, leading to poor identifiability of individual parameters despite excellent agreement with data. Consequently, conventional uncertainty quantification, sensitivity analysis, and parameter estimation can become unstable or misleading, and prior assumptions may dominate posterior inferences. CRISP reframes model analysis around preserving information for decision-relevant predictions by targeting the few stiff directions controlling clinically meaningful outcomes.

Conclusion: CRISP provides a unifying, cross-domain framework for reasoning about large PBPK models. By treating sloppiness as a structural feature rather than a modeling failure, this approach supports more credible, auditable, and interpretable use of complex mechanistic models as they continue to grow in scale and scope.

CONTACT

Mark K. Transtrum, PhD
Cross Stream Bioanalytics
1519 S Roberta St, Salt Lake City UT 84115
mark@cross-stream.ai
+1(760) 712-9009

INTRODUCTION

Models can fit data from patient's biodistribution almost perfectly and still tell you almost nothing reliable about the what actually matters for the next treatment cycle: the cumulative absorbed dose in the tumor and off-target organs. This is a consequence of parameters constrained by fit-to-scan-data are not the same as the parameters which drive AUC dose.

Much of the AI brought to bear on this gap works by training a black-box model and then trying to explain, after the fact, what it did — a strategy increasingly recognized as treating the symptom rather than the disease (Rudin, 2019). The alternative is to build models that are interpretable from the start. The mathematical machinery for doing this rigorously is called information geometry; it has been established for decades (Amari, 2016) and specifically diagnoses the gap between the parameters constrained by experiment and those needed for clinical decisions (Transtrum, Machta & Sethna, 2010).

CRISP is the first software that repairs this gap. Here, applied to a validated compartmental dosimetry model (Kletting et al., 2016), it reveals that the parameters best constrained by a patient's scan are often not the parameters that determine their AUC, a failure mode invisible to standard sensitivity analysis. By contrast, it provides real personalized dosimetry, on a laptop, in minutes. The method generalizes well to other problems in the biomedical space.

METHODS AND MATERIALS

Data: serial activity measurements (planar γ-camera and serum) following 90Y-DOTATATE administration were as previously described (Kletting et al., 2016), collected up to 9 time points per compartment across blood, kidney, liver, spleen, whole-body, and two tumor regions (n=39 observations per data set). **Model:** A whole-body 16-compartment ODE (adapted from Kletting et al. 2016), comprising 130 state variables and 179 parameters (65 free). **Fit** to the activity time-courses with self-authored CRISP-based software; error was essentially zero. **Sensitivity and Predictive Queries:** we computed the Fisher Information Matrix (FIM) for two distinct observation sets: (1) the calibration data itself (7 compartments at 9 times); and (2) model-forecast AUC in kidney, liver, spleen, and/or red marrow at 7d post injection. Eigendecomposition of each FIM identifies the parameter combinations best constrained by that observation set. Comparing eigenvector structure between the two FIMs tests directly whether the parameters a clinician's scan data actually pin down are the same ones that determine the clinically relevant AUC — the central question of this work. Full CRISP workup involving relevons and identifying a supremum were not done here.

RESULTS

Modeling scan activity: For each parameter, we calculated Cramer-Rao importance score. This statistic revealed a striking sparsity: just $S_i = \sum_k \lambda_k v_{ki}^2$ model's parameters account for >95% of total model. The single most important parameter, F_Mus, controls ~40% of output. **Predictive query:** When predicting exposure at 7d (AUC in 4 organs), eigenanalysis of the FIM reveal F_Mus was again a dominant parameter; constrained by imaging and clinically decisive it is a good biomarker. By contrast, kPR — a parameter that is both easily measurable and highly variable across patients — accounts for ~20% of total scan sensitivity but <1% of AUC sensitivity. This demonstrates an important failure mode: a parameter can dominate how well a model fits imaging data while being almost irrelevant to the clinical quantity that data is meant to inform, a distinction invisible to fit-quality metrics alone. Predictive queries without the Red Marrow AUC simply removed PS_RM as predictive (not shown).

Fig 1. Sensitivity of activity measurements to model parameters

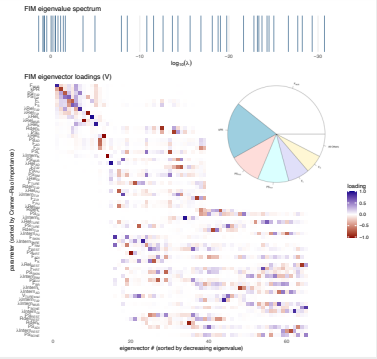
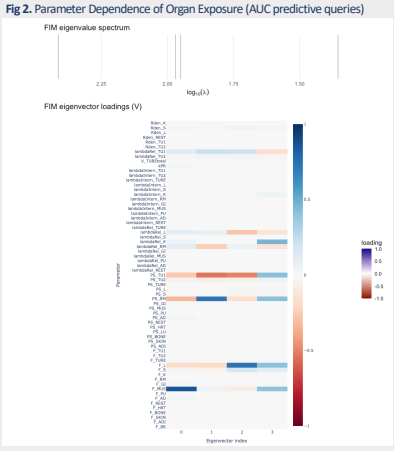


Fig 2. Parameter Dependence of Organ Exposure (AUC predictive queries)



DISCUSSION

This dissociation between scan-constrained and clinically-relevant parameter sensitivity is a specific, clinically consequential instance of a more general phenomenon documented across systems biology and physics: complex multiparameter models are typically "sloppy," with behavior governed by a small number of stiff parameter combinations while the rest remain only loosely constrained by any single dataset (Transtrum, Machta & Sethna, 2010). What this analysis adds is that sloppiness is not merely a nuisance for parameter identifiability — it is directional, and the stiff directions relevant to fitting scan data need not coincide with the stiff directions relevant to a downstream clinical quantity like AUC. The red marrow result reinforces this: removing an unmeasured compartment did not simply redistribute uncertainty among remaining parameters, it eliminated an entire sensitivity mode. A model's effective dimensionality is contingent on the question posed, not an intrinsic property of the equations. The broader critique is that post-hoc explanation of a fitted black-box model risks answering the wrong question entirely (Rudin, 2019); even a fully mechanistic, interpretable-by-design ODE model can mislead if its fit quality brings over-confidence in an unrelated output.

CONCLUSIONS

These results argue for treating parameter importance as query-specific rather than fixed, and for evaluating it directly rather than assuming it. CRISP makes visible failure modes like the F_Mus/kPR dissociation that a raw goodness-of-fit metric cannot reveal. Applied prospectively, this means a clinician could fit a patient's scan data, then directly query the model for the AUC-driving parameters. The full CRISP pipeline, including relevon and supremum-based model reduction, extends this further: producing not just a ranked sensitivity but a maximally reduced, patient-specific model tailored to the clinical question being asked. This capability is implemented in self-authored software, currently available for beta testing.

REFERENCES

1. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. DOI:10.1038/s42256-019-0048-x
2. Amari, S. (2016). *Information Geometry and Its Applications. Applied Mathematical Sciences*, vol. 194. Springer.
3. Transtrum, M. K., Machta, B. B., & Sethna, J. P. (2010). Why are nonlinear fits to data so challenging? Physical Review Letters, 104(6), 060201. DOI:10.1103/PhysRevLett.104.060201
4. Kletting, P., (5 authors) & Glatting, G. (2016). Optimized peptide amount and activity for 90Y-labeled DOTATATE therapy. Journal of Nuclear Medicine, 57(4), 503–508. DOI:10.2967/jnumed.115.164699