Applications of unsupervised learning for plasma concentration-time curves

BASS 2024, November 6

Junghi Kim

U.S. Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Office of Translational Sciences, Office of Biostatistics

Disclosures

Conflict of Interest

The present study reflects the views of the authors and should not be construed to represent the views or recommendations of the U.S. Food and Drug Administration.

Funds

This project was supported in part by an appointment to the Research Participation Program at the FDA by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy and the FDA.

Acknowledgement

Jackson Lautier, Bentley University

Stella Grosser, CDER/Office of Biostatistics

Jessica Kim, CDER/Office of Biostatistics

Hyewon Kim, CDER/Office of Clinical Pharmacology

Outline

AL/ML in Regulatory Science and Background

Existing Method for Clustering PK Curves

Proposed Method

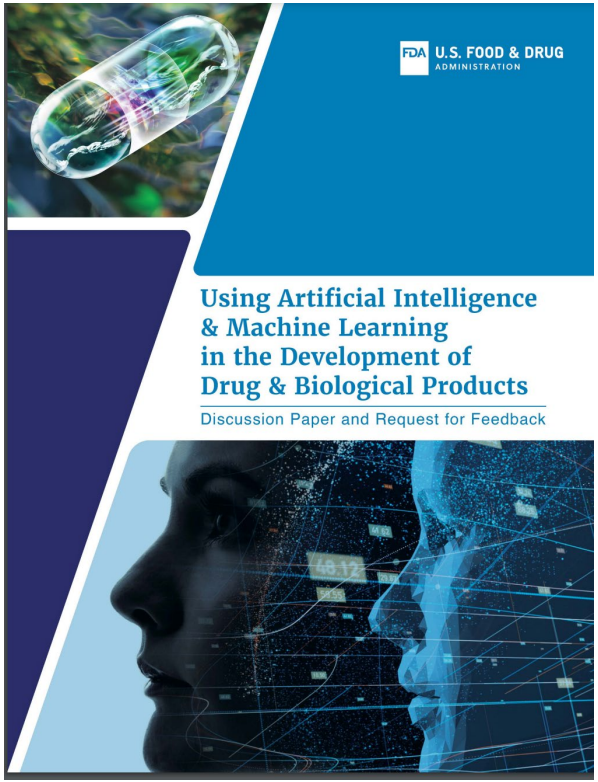
Simulation Studies

Application

Discussion

AL/ML in Regulatory Science

Discussion Paper from May 2023



Joint effort of CDER, CBER, CDRH and the Digital Health Center of Excellence (DHCoE)

Landscape of current and potential uses of AI/ML

Considerations for the use of AI/ML

Next steps and stakeholder engagement

Landscape Analysis for AI/ML Related Submissions at CDER/FDA

PERSPECTIVES

PERSPECTIVE

Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021

Qi Liu[1,†], Ruihao Huang[1,†], Julie Hsieh[1,†], Hao Zhu[1,*,†], Mo Tiwari[1], Guansheng Liu[1], Daphney Jean[1], M. Khair ElZarrad[2], Tala Fakhouri[2], Steven Berman[3], Billy Dunn[3], Matthew C. Diamond[4] and Shiew-Mei Huang[1]

An analysis of regulatory submissions of drug and biological products to the US Food and Drug Administration from 2016 to 2021 demonstrated an increasing number of submissions that included artificial intelligence/machine learning (AI/ML). AI/ML was used to perform a variety of tasks, such as informing drug discovery/repurposing, enhancing clinical trial design elements, dose optimization, enhancing adherence to drug regimen, endpoint/biomarker assessment, and postmarketing surveillance. AI/ML is being increasingly explored to facilitate drug development.

BACKGROUND

Over the past decade, there has been a rapid expansion of artificial intelligence/machine learning (AI/ML) applications in biomedical research and therapeutic development. In 2019, Liu *et al*, provided an overview of how AI/ML was used to support drug development and regulatory submissions to the US Food and Drug Administration (FDA). The authors envisioned that AI/ML would play an increasingly important role in drug development.[1] That prediction has now been confirmed by this landscape analysis based on drug and biologic regulatory submissions to the FDA from 2016 to 2021.

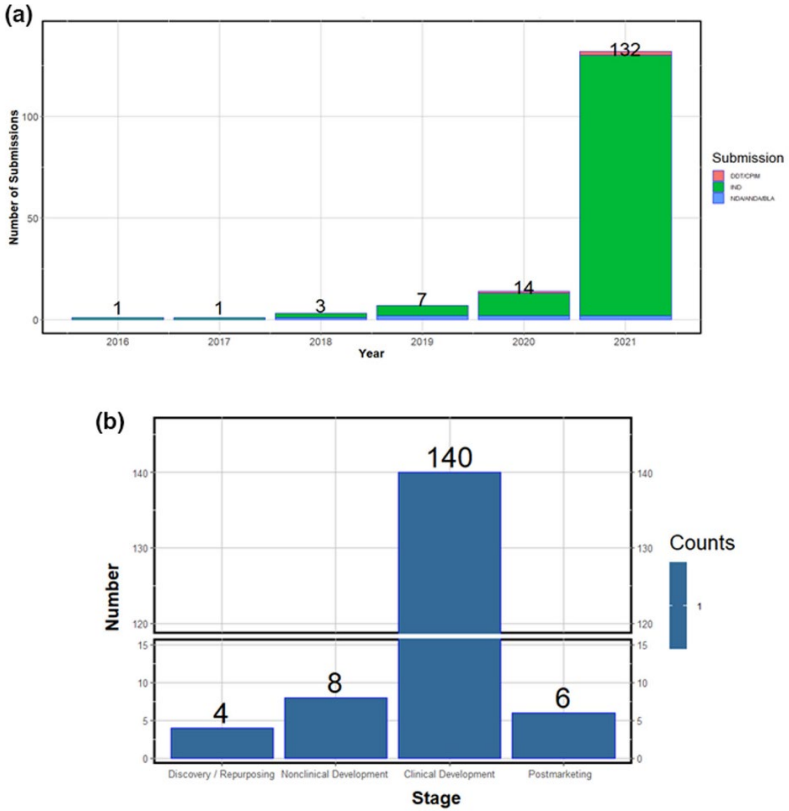
THE TREND OF INCREASING AI/ML–RELATED SUBMISSIONS AT THE FDA'S CENTER FOR DRUG EVALUATION AND RESEARCH

This analysis was performed by searching for submissions with key terms "machine learning" or "artificial intelligence" in Center for Drug Evaluation and Research (CDER) internal databases for Investigational New Drug applications, New Drug Applications, Abbreviated New Drug Applications, and Biologic License Applications, as well as submissions for Critical Path Innovation Meeting and the Drug Development Tools Program. We evaluated all data from 2016 to 2021. Figure 1a demonstrates that submissions with AI/ML components have increased rapidly in the past few years. In 2016 and 2017, we identified only one such submission each year. From 2017 to 2020, the numbers of submissions increased by approximately twofold to threefold yearly. Then in 2021, the number of submissions increased sharply to 132 (approximately 10-fold as compared with that in 2020). This trend of increasing submissions with AI/ML components is consistent with our expectation based on the observed increasing collaborations between the pharmacentrical and technology industries.

Figure 1b illustrates the distributions of these submissions by therapeutic area. Oncology, psychiatry, gastroenterology, and neurology were

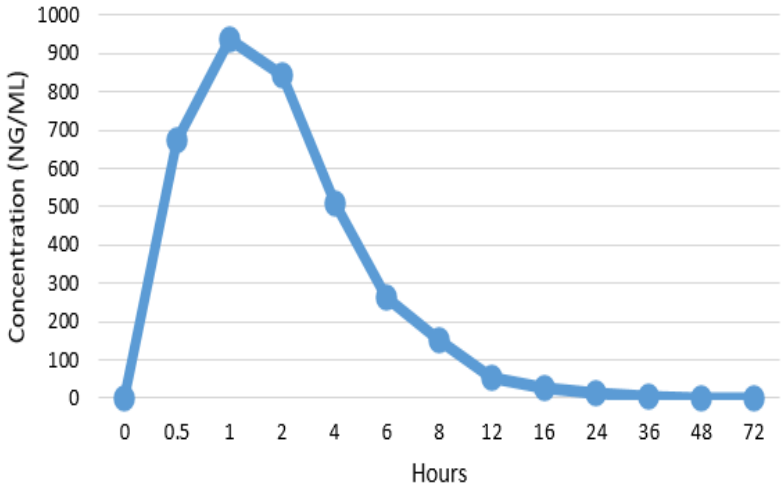
[1]Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [2]Office of Medical Policy, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [3]Office of New Drugs, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [4]Digital Health Center of Excellence, Center for Devices and Radiological Health (CDRH), US Food and Drug Administration, Silver Spring, Maryland, USA.
*Correspondence: Hao Zhu (hao.zhu@fda.hhs.gov)
†These authors contributed equally.

Received March 16, 2022; accepted May 19, 2022. doi:10.1002/cpt.2668



Background

Pharmacokinetic Concentration Curve



Plasma concentration time curves or pharmacokinetic (PK) curves are generated by plotting drug concentration levels in plasma samples at various time intervals after the administration of a drug product (Shargel and Yu, 2016).

Pharmacokinetic Parameters

Compartment model

a model-based method employing a mathematical equation describing the concentration of drug in the body at any given time

i.e., absorption rate (κ_a)and elimination rate (K)

Non compartment model:

a model-independent method, where PK parameters are estimated directly from the observed data (PK curves) requiring no historical knowledge of the PK characteristics of the drug in the body

i.e., Maximum concentration (Cmax), Time of maximum concentration (Tmax), Area under the concentration-time curve (AUC)

Why Cluster Pharmacokinetic (PK) Curves?

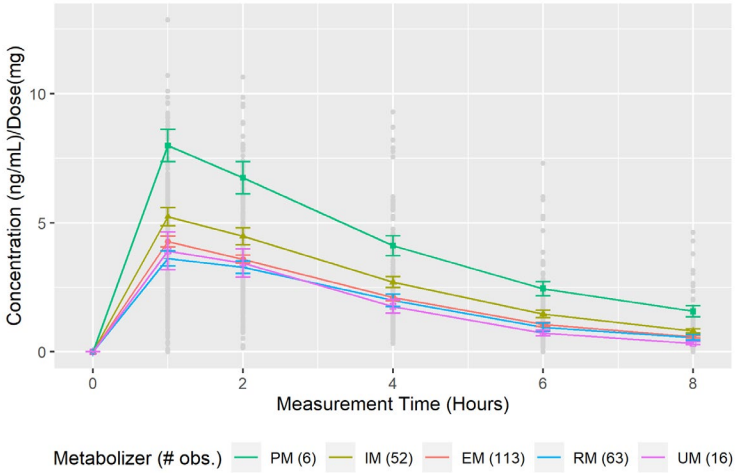


Figure: PK Curves by Subject Metabolic Rate

Grouping patients by PK concentration curve shape may help researchers identify patterns among subjects, which may inform dosages or therapeutic strategies.

Existing Methods

How to Cluster PK Curves

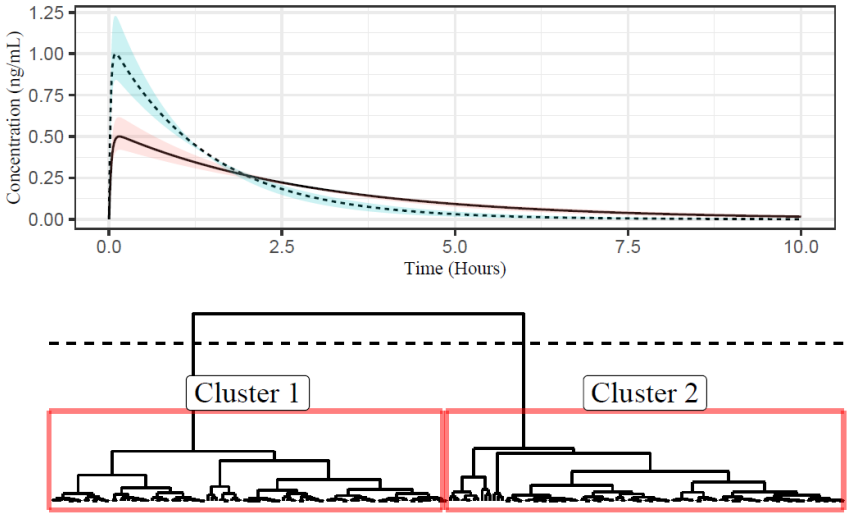
A key issue in clustering is the choice of a suitable measure to assess the distance between two-time course data and the corresponding clustering method.

Treat PK curves as time series data objects that may be characterized by geometric shape and clustered with hierarchical clustering (Lautier et al. 2024).

Despite a robust time-series data object clustering literature (e.g., Montero and Vilar 2014), there is no dissimilarity measure tailored for the PK curves.

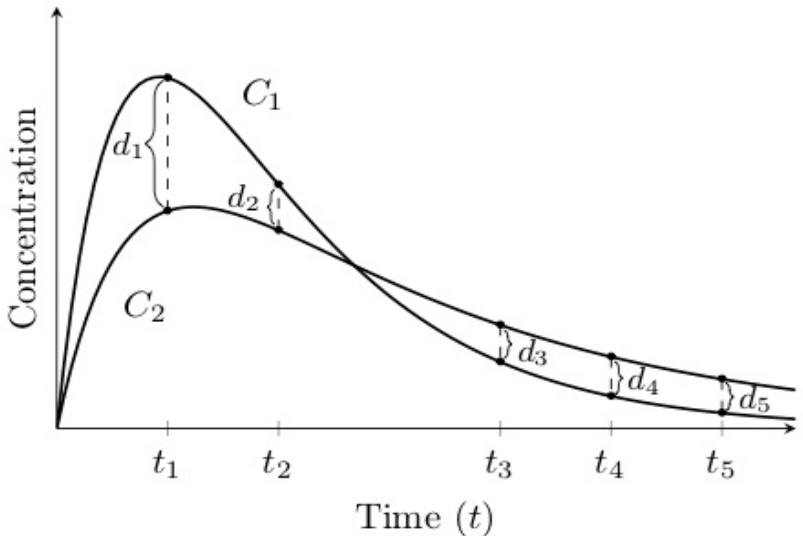
Hierarchical Clustering

Unsupervised Clustering is a machine learning technique in which similar objects are algorithmically divided into distinct groups.



Dendrograms can help decide on the appropriate number of clusters (Lautier et al. 2024).

What is PK Curve Dissimilarity?



e.g., Euclidian

$$d_{L_2(Curve_1,Curve_2)} = \sqrt{\sum_{\{i=1\}}^{5} d_i^2}$$

Dissimilarity Measures

- Of the many time series dissimilarity measures (i.e., distance), five are theoretically reasonable for PK curves (Lautier et al. 2024, Montero and Vilar 2014):
  [1] Euclidean distance (shape)
  [2] Dynamic Time-Warping (shape)
  [3] Fr´echet Distance (shape)
  [4] Correlation-based (structure)
  [5] temporal correlation coefficient (shape-structure)

- Of these five, Euclidean distance offered the best combination of performance, theoretical justification, and ease of interpretation for presumed end-users (pharmacologists).

Limitations of Existing Methods

- For applying Euclidian/correlation-based distances, only the common sampling time points across PK curves can be used for a clustering task

- Often the sampling time points of PK concentration are not consistent across studies. Therefore, without any adjustments (or interpolation) to the data, only the distance metrics Fréchet and DTW could be applied

- The Fréchet and DTW distance metrics are likely to distort time between concentration sampling points between PK curves

Proposed Method

Proposed Dissimilarity for PK Curve

Goal: propose a geometric-shape-based dissimilarity measure
that is (1) informed by a PK curves and (2) more robust to situations with high clustering difficulty

PK curves are often summarized numerically by PK parameters; e.g.,

- Area-Under-the-Curve (AUC)
- Maximum Concentration (CMAX)
- Time-until-Maximum-Concentration (TMAX)

The PK parameters (AUC, CMAX, TMAX) are relied upon by pharmacologists to analyze and study drug response, and they (1) do not assume any underlying structural model and (2) may be estimated if sampling time points differ between PK curves.

Proposed Dissimilarity for PK Curve

Let p_y be an l-dimensional vector of PK parameters corresponding to a single PK curve y. That is, $p_y = \left(p_{1y}, \ldots, p_{1y}\right)' \in \mathrm{R}^l$. Similarly, let $p_x = (p_{1x}, \ldots, p_{1x})' \in \mathrm{R}^l$ correspond to a second PK curve x. Define l positive weights $0 \leq w_i \leq 1,\ i = 1, \ldots, l$, such that $\sum_i w_i = 1$. We define an importance-weighted distance as

$$d^w(p_y, p_x) = \left(\sum_i^l w_i (p_{iy} - p_{ix})^2 \right)^{\frac{1}{2}}$$

The above distance function is a valid metric or distance function which satisfies Rudin (1976, Definition 2.15, pg. 30).

How to determine the weights $w_i, 1 \leq i \leq l$?

- **One Approach**

Rely on pharmacologist expertise to determine which PK parameters should have the most importance (i.e., weight)

- **Second Approach**

Allow data to inform weights

1) Multimodality in a density plot of each PK parameter across a data set may indicate possible clusters.
2) Hence, assign weights $w_i, 1 \leq i \leq l$, for each PK parameter in the distance calculation by more heavily weighting PK parameters that exhibit greater relative multimodality.
3) Hartigan and Hartigan (1985) propose a "dip test" of unimodality in a general setting, which can be used as an automated way to assign weights.

Visualizing ‘Dip Test’ of Hartigan and Hartigan (1985)

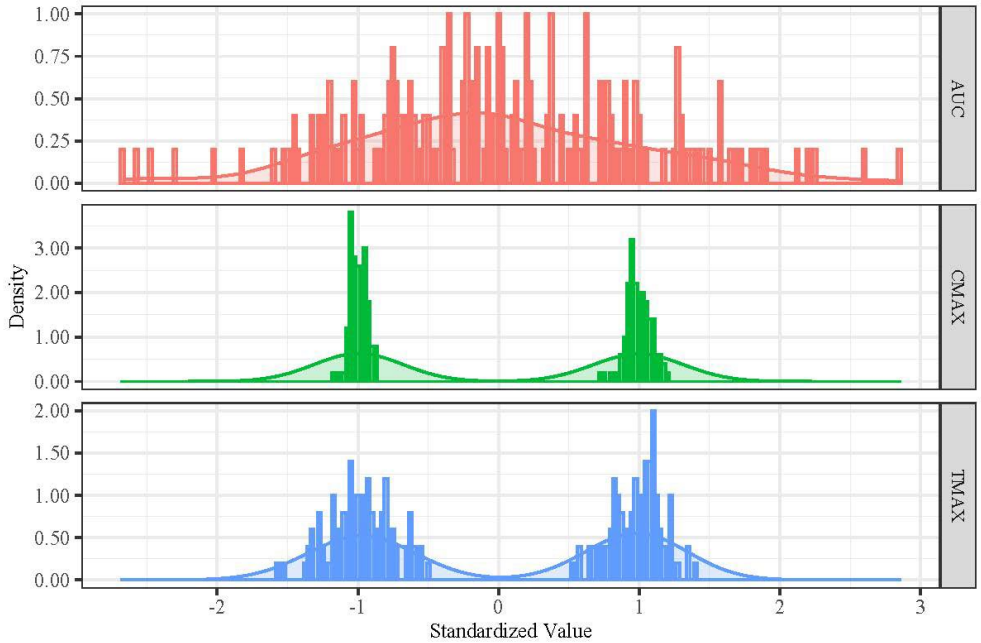


Figure: Empirical Density Plots of Two Simulated Groups

Visualizing ‘Dip Test’ of Hartigan and Hartigan (1985)

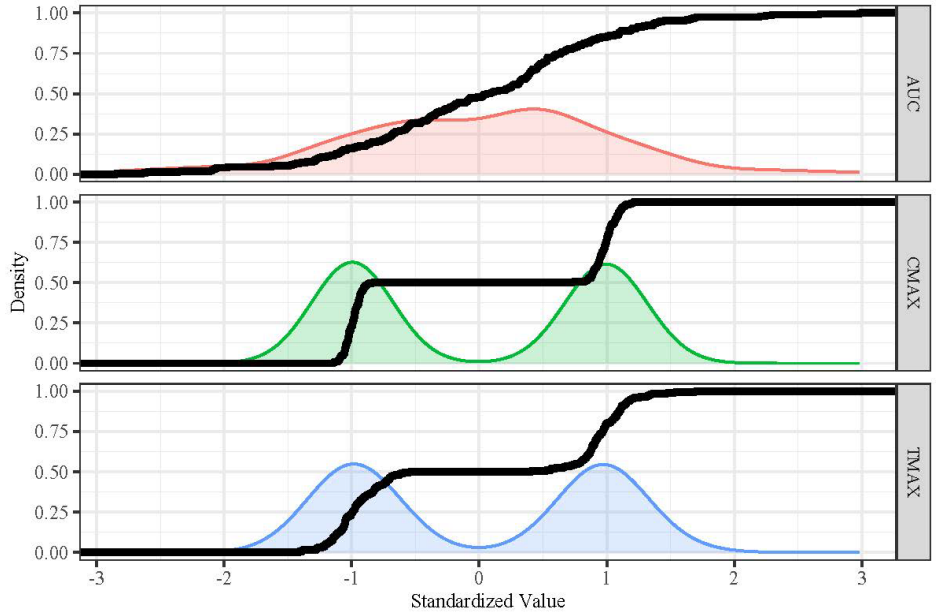


Figure: Empirical Density and CDF Plots of Two Simulated Groups

pkDip Distance Algorithm

1. Select l PK parameters for all observations in data set.
2. For each PK parameter selected in, center and standardize the observations. If necessary, apply a continuity correction[1].
3. For each standardized PK parameter in (2), perform the "dip test" of Hartigan and Hartigan (1985) and calculate the p-value. Denote this value $p_i, i = 1, \ldots, l$
4. Calculate

$$w_i = \frac{1/p_i}{\sum_{k=1}^{l} 1/p_k}, \qquad i = 1, \ldots, l$$

5. With $w_i, i = 1, \ldots, l$ from (4), calculate $d^w\left(p_j, p_{j'}\right)$ for all $1 \leq j, j' \leq n,\ j = j'$. These relative distances may then be used to perform hierarchical clustering.

[1]Hartigan and Hartigan (1985) assume continuous data, whereas TMAX is often discrete data.

Simulation Studies

Simulation Scenarios

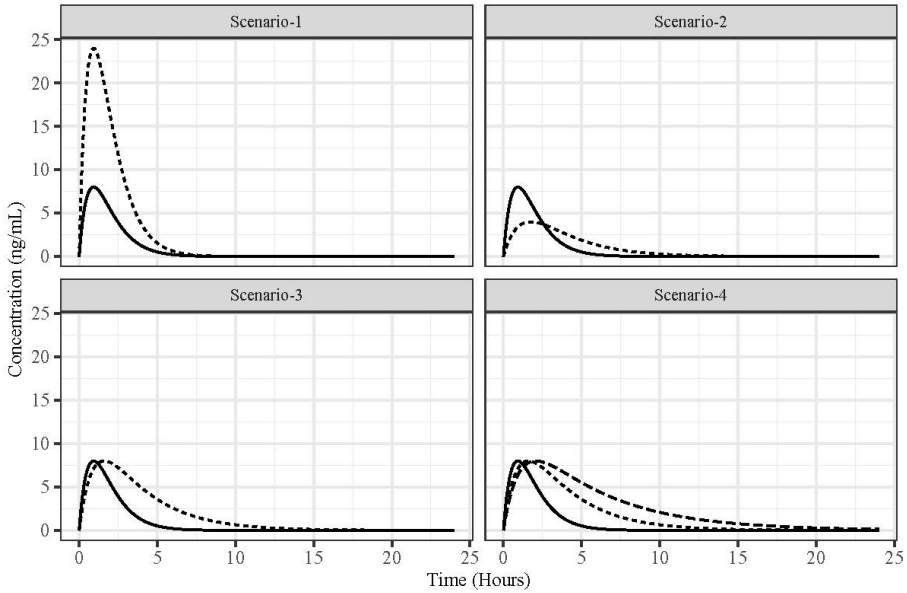


Figure: Representative Mean PK Curve of Simulated Clusters

Notes: All four Scenarios are evaluated with varying clustering difficulty and number of measurement time points (full vs. sparse). Scenario 3 is evaluated again for unbalanced sample size (i.e., Scenario 3').

Simulation Scenarios

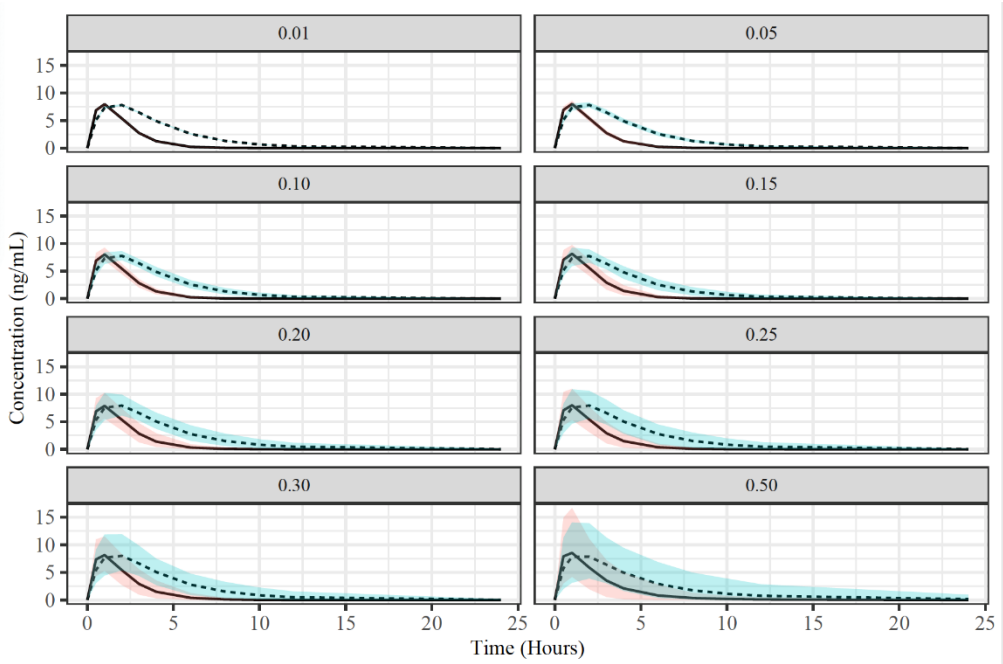


Figure: Clustering Difficulty Illustration: $\alpha = \{0.01, .. 0.5\}$

Notes: The mean of the two clusters follows the solid and dashed lines, respectively, and the shaded ribbons represent 95% empirical confidence intervals as the within cluster variability increases

Simulation Study Scenario Summary

- Each cluster consists of 100 curves, and each scenario includes 200 replicates
- PK curves are generated from one compartment linear model ($K_a, K\ and\ V_c$)
- The mean of each parameter is selected to generate desired AUC, C_{max}, or T_{max}
- Assume assume each of the parameters,

$$K_a \sim log\ N \text{ with } E(K_a) = \rho \text{ and } SD(K_a) = \alpha\rho$$

- Clustering difficulty ranging across

  α = (0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.50)'

- Full/Sparse time points are

  t_{full} = (0, 0.5, 1, 2, 3, 4, 6, 8, 10, 12, 24, 36, 48)'

  t_{sparse} = (0, 0.5, 1, 2, 4, 8)'

Simulation Results Overview

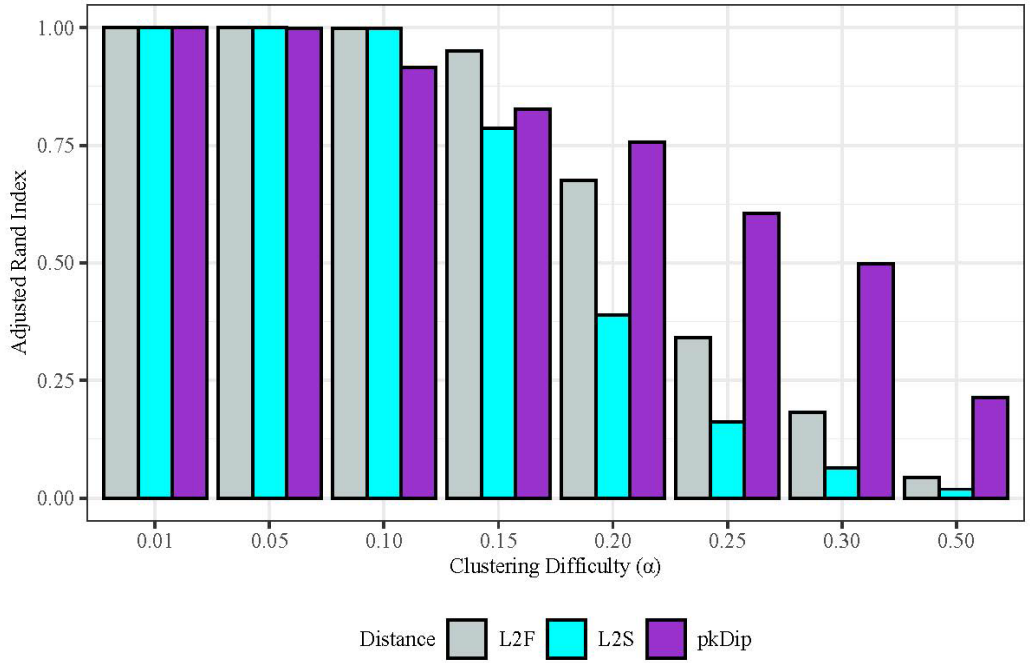


Figure: Simulation Study Results (Adjusted Rand Index) for a scenario with equal C_{max}, different AUCs, 2 Clusters, Unbalanced sample size

Simulation Results Summary

- We evaluated clustering performance with varying clustering difficulty, number of measurement timepoints and sample size of each cluster
- The pkDip distance algorithm obtains comparable or better clustering performance to L^2 when a clustering task is easy
- The pkDip distance algorithm underperforms L^2 for some scenarios in moderate clustering difficultly ($\alpha \in \{0.10, 0.15\}$).
- The pkDip distance algorithm is noticeably more robust:
  - (1) better performance than L^2 for high clustering difficulty (α>0.20)
  - (2) better performance than L^2 for high clustering difficulty (α>0.20) and sparse measurement times
  - (3) better performance than L^2 for unbalanced clusters.

Application

Case Study

- Background

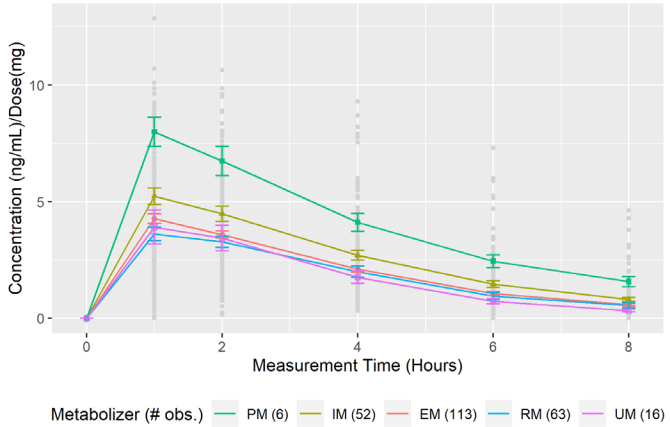
    1) Genetic factors are known to cause inter-individual differences in pharmacokinetics (Ahmed et al. 2016)

    2) FDA recommends a reduced dosage for patients classified as a CYP2C19 poor metabolizer (PM) in many disease areas (FDA, 2021)

- Our underlying data includes nine Phase 1 studies in which we have subjects' genetic information of CYP2C19 and CP2C9, as well as PK concentration curves

- In this study, it was shown that AUC of poor metabolizers (PM) of CYP2C19 was 2 or 3-fold increased compared to other metabolizers. It was suggested to use a dosage reduction for PM patients.

Summery Statistics for Case Study Data



CYP		Number of	C_{max} (ng/ML)/mg	AUC_{last} (ng*hr/ML)/mg
Phenotype	Genotype	Observations	Mean (Std)	Mean (Std)
PM	*2/*2	6	9.03 (1.64)	42.09 (8.64)
IM	*1/*2	35	6.21 (2.69)	25.64 (11.18)
IM	*2/*17	17	6.03 (2.17)	26.49 (9.39)
EM	*1/*1	113	5.44 (2.45)	20.06 (9.63)
RM	*1/*17	63	4.88 (2.42)	18.49 (12.30)
UM	*17/*17	16	4.29 (2.68)	15.29 (8.77)
Total Sample		**250**	**5.46 (2.55)**	**21.11 (11.32)**

Goal of the Case Study

Can we independently recover the same potential of increased drug exposure for PMs by using unsupervised ML on only subject PK curve data

To this end, we connect the clustering labels with metabolizer status and PK parameters

Clustering Difficulty Estimate in Case Study Data

	Coefficient of Variation (CV)			
α	k_a	$\mathcal{K}$	V_c	Row-wise CV Mean
0.01	0.3859	0.3820	0.0672	0.2784
0.05	0.3868	0.3880	0.0800	0.2849
0.10	0.4111	0.3950	0.1070	0.3044
0.15	0.4899	0.4284	0.1559	0.3581
0.20	0.5467	0.4085	0.1753	0.3768
0.25	0.6105	0.4580	0.2202	0.4295
0.30	0.6659	0.4557	0.2468	0.4561
0.50	0.7228	0.6330	0.4382	0.5980
Data	0.4441	0.5656	0.6070	0.5389

Table: Coefficient of variation of PK parameters of one compartment model

Case Study Data

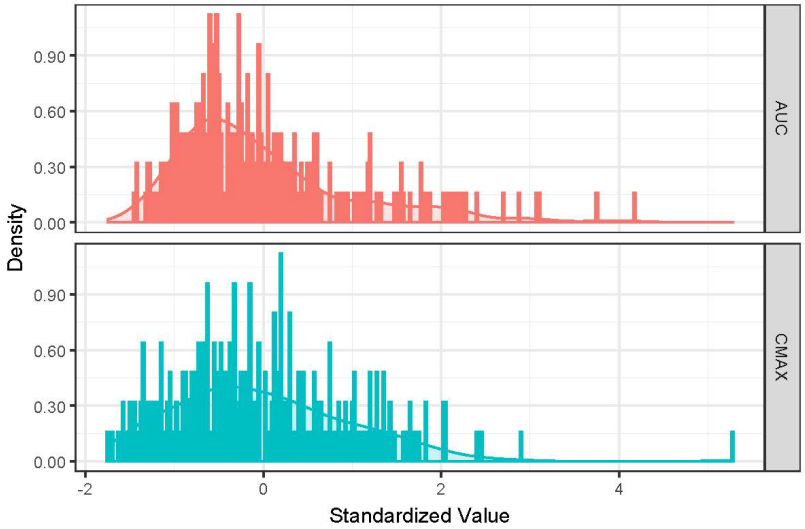


Figure: Empirical Density Plots by PK Parameter (n=250)

(1) Tmax was not used to mirror the original reference pharmacogenomic analysis
(2) The PK parameter weights for pkDip are 0.47 for AUC and 0.53 for Cmax

Clustering Results

Euclidean Distance (Lautier et al. 2024)

Cluster	1	2	3	4	**Total**
# Obs. (%)	204 (81.6%)	39 (15.6%)	3 (1.2%)	4 (1.6%)	**250**
Distribution of Metabolizer Status by Cluster					
PM	1 (16.67%)	5 (83.33%)	0 (0%)	0 (0%)	**6 (100%)**
IM	35 (67.31%)	16 (30.77%)	0 (0%)	1 (1.92%)	**52 (100%)**
EM	99 (87.61%)	11 (9.73%)	2 (1.77%)	1 (0.88%)	**113 (100%)**
RM	55 (87.30%)	5 (7.94%)	1 (1.59%)	2 (3.17%)	**63 (100%)**
UM	14 (87.50%)	2 (12.50%)	0 (0%)	0 (0%)	**16 (100%)**
Average (Standard deviation) PK Metrics by Cluster					
AUC_{last}	17.12 (6.76)	36.27 (8.28)	60.01 (10.75)	47.10 (10.69)	**21.11 (11.32)**
C_{max}	4.66 (1.89)	9.04 (2.17)	10.22 (0.45)	7.48 (1.16)	**5.46 (2.55)**

pkDip Distance

Cluster	1	2	3	4	**Total**
# Obs. (%)	180 (72.0%)	67 (26.8%)	2 (0.8%)	1 (0.4%)	**250**
Distribution of Metabolizer Status by Cluster					
PM	0 (0%)	6 (100%)	0 (0%)	0 (0%)	**6 (100%)**
IM	29 (55.77%)	23 (44.24%)	0 (0%)	0 (0%)	**52 (100%)**
EM	87 (76.99%)	20 (20.35%)	2 (1.77%)	1 (0.88%)	**113 (100%)**
RM	50 (79.37%)	13 (20.63%)	0 (0%)	0 (0%)	**63 (100%)**
UM	14 (87.50%)	2 (12.50%)	0 (0%)	0 (0%)	**16 (100%)**
Average (Standard deviation) PK Metrics by Cluster					
AUC_{last}	15.85 (5.49)	33.61 (10.08)	65.95 (3.45)	38.99 (0.00)	**21.11 (11.32)**
C_{max}	4.24 (1.52)	8.40 (1.42)	10.43 (0.39)	18.95 (0.00)	**5.46 (2.55)**

Summary of Clustering Results

- Our work appears to correspond to the results of Euclidian distance in that all six Poor Metabolizers find the same cluster (Cluster 2).
- Lautier et al. 2024 identify IMs as potentially worthy of further study regarding a reduced dosage.
- Our results with pkDip appear to warrant further evidence that IMs may be worthy of additional study.

Discussion

Conclusions

- At present, there are no geometric-shape-based distances specifically designed for PK curves with hierarchical clustering though Lautier et al. (2024) find that L^2 distance can perform well in ideal data conditions.
- We propose the pkDip distance algorithm, which dynamically weights PK parameters by multimodality.
- Similar performance to L^2 distance in ideal data conditions and appears more robust to clustering difficulty, sparse measurement times, and unbalanced clusters: all likely in PK curve data.
- The pkDip distance algorithm also offers an ease of interpretation & flexibility to pharmacologists.
- Applications outside of PK curves?

References

Montero, P. and Vilar, J.A. (2014). TSclust: An R Package for Time Series Clustering. Journal of Statistical Software, 62(1), 1–43.

Ahmed, S., Zhou, Z., Zhou, J., and Chen, S. Q. (2016). Pharmacogenomics of Drug Metabolizing Enzymes and Transporters: Relevance to Precision Medicine. Genomics, proteomics & bioinformatics, 14(5), 298–313.

Hartigan, J.A. and Hartigan, P.M. (1985). The dip test of unimodality. The Annals of Statistics 13, 70–84.

Lautier J.P., Grosser, S., Kim, J., Kim, H., and Kim, J. (2024). Clustering plasma concentration-time curves: applications of unsupervised learning in pharmacogenomics. Journal of Biopharmaceutical Statistics pp. 1–19.

Rudin, W. (1976). Principles of Mathematical Analysis. McGraw-Hill, Inc.

Tom'as, E. and Vinga, S. (2017). Unsupervised learning of pharmacokinetic responses. Computational Statistics 32, 409–428.

Shargel, L. and Yu, A. (2016). Applied Biopharmaceutics & Pharmacokinetics, Seventh Edition McGraw Hill


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Back up slides

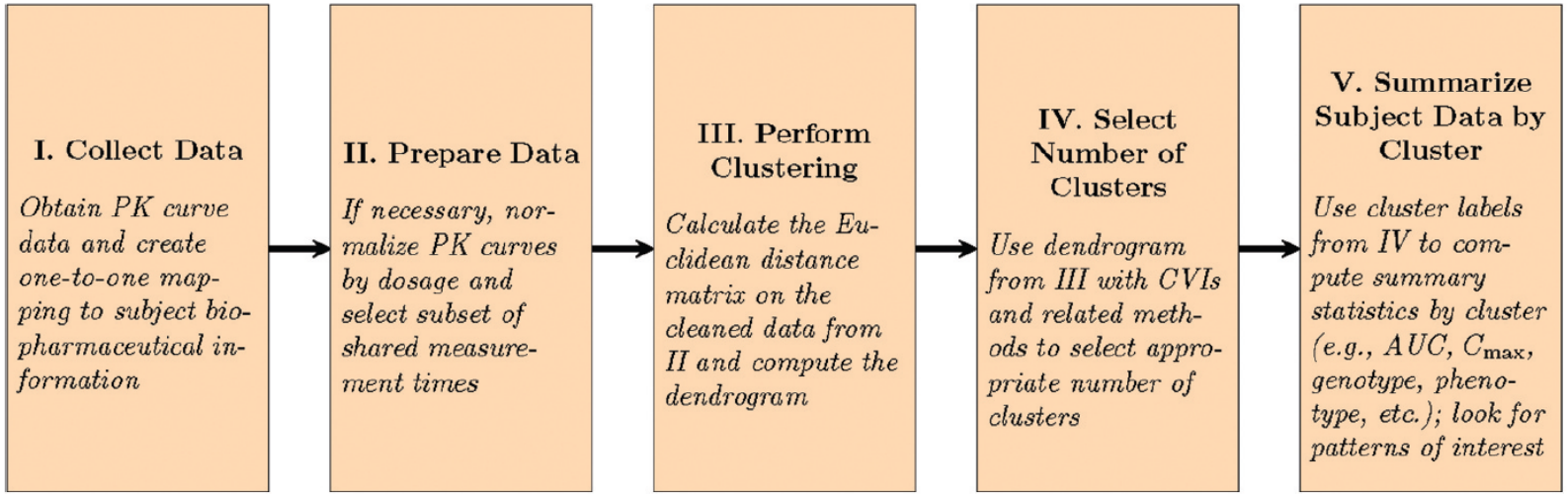


Figure: Our case study process is summarized

How to Simulate Data

Consider two PK curves, C1 and C2, from a one compartment linear PK model, assuming first-order absorption and first-order elimination after oral administration. Specifically, we assume the first dosage is administered at $t = 0$. That is

$$C(t) = \frac{DK_a}{V_c(k_a - k_{el})}\{\exp(-k_{el}t) - \exp(-k_a t)\}$$

where D is the dosage, V_c is the central volume, k_a is the first order absorption, and $k_{el} = \mathrm{K}/V_c$ (where K is the clearance).

By adjusting K, V_c and k_a, we generate wanted AUC and Cmax.

Simulation Results

		t_{full}					t_{sparse}			
Scen.	α	L_2	CORT	COR	DTW	**pkDip**	L_2	CORT	COR	DTW
1	0.01	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)
	0.05	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)
	0.10	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)
	0.15	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.95 (0.21)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.98 (0.14)
	0.20	0.92 (0.25)	0.87 (0.33)	0.00 (0.00)	0.74 (0.43)	**0.94 (0.22)**	0.93 (0.24)	0.90 (0.29)	0.00 (0.00)	0.76 (0.41)
	0.25	0.73 (0.43)	0.63 (0.47)	0.00 (0.00)	0.45 (0.48)	**0.79 (0.37)**	0.75 (0.41)	0.69 (0.44)	0.00 (0.00)	0.45 (0.47)
	0.30	0.32 (0.44)	0.22 (0.39)	0.00 (0.00)	0.11 (0.27)	**0.34 (0.43)**	0.36 (0.44)	0.31 (0.43)	0.00 (0.00)	0.14 (0.29)
	0.50	0.02 (0.04)	0.01 (0.04)	0.00 (0.00)	0.02 (0.06)	**0.00 (0.00)**	0.02 (0.05)	0.02 (0.07)	0.00 (0.00)	0.02 (0.04)
2	0.01	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	0.05	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	0.10	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)	1.00 (0.00)	**0.97 (0.04)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	0.15	0.99 (0.02)	0.99 (0.01)	0.96 (0.04)	0.96 (0.17)	**0.75 (0.30)**	0.99 (0.02)	0.99 (0.02)	0.96 (0.04)	0.98 (0.10)
	0.20	0.78 (0.35)	0.93 (0.11)	0.83 (0.15)	0.64 (0.42)	**0.54 (0.32)**	0.74 (0.38)	0.92 (0.14)	0.80 (0.21)	0.69 (0.41)
	0.25	0.37 (0.41)	0.73 (0.29)	0.63 (0.23)	0.30 (0.37)	**0.45 (0.28)**	0.41 (0.41)	0.73 (0.30)	0.54 (0.31)	0.32 (0.39)
	0.30	0.12 (0.25)	0.48 (0.34)	0.46 (0.26)	0.10 (0.20)	**0.27 (0.26)**	0.14 (0.25)	0.46 (0.35)	0.37 (0.27)	0.11 (0.21)
	0.50	0.01 (0.02)	0.04 (0.10)	0.11 (0.14)	0.02 (0.05)	**0.02 (0.05)**	0.01 (0.03)	0.04 (0.09)	0.12 (0.12)	0.01 (0.03)
3	0.01	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	0.05	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.98 (0.10)	**0.98 (0.02)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)
	0.10	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.31 (0.35)	**0.80 (0.06)**	0.99 (0.07)	1.00 (0.01)	1.00 (0.01)	0.45 (0.43)
	0.15	0.92 (0.21)	0.96 (0.08)	0.95 (0.05)	0.04 (0.09)	**0.64 (0.14)**	0.70 (0.43)	0.90 (0.23)	0.93 (0.11)	0.04 (0.11)
	0.20	0.56 (0.41)	0.80 (0.20)	0.79 (0.15)	0.02 (0.04)	**0.45 (0.23)**	0.27 (0.40)	0.66 (0.35)	0.76 (0.19)	0.01 (0.03)
	0.25	0.18 (0.31)	0.49 (0.33)	0.57 (0.22)	0.01 (0.02)	**0.39 (0.21)**	0.06 (0.19)	0.33 (0.35)	0.50 (0.27)	0.01 (0.01)
	0.30	0.04 (0.13)	0.24 (0.28)	0.38 (0.25)	0.01 (0.01)	**0.32 (0.19)**	0.01 (0.02)	0.15 (0.24)	0.33 (0.25)	0.00 (0.01)
	0.50	0.00 (0.00)	0.02 (0.07)	0.09 (0.12)	0.00 (0.00)	**0.03 (0.07)**	0.00 (0.00)	0.01 (0.04)	0.11 (0.11)	0.00 (0.00)

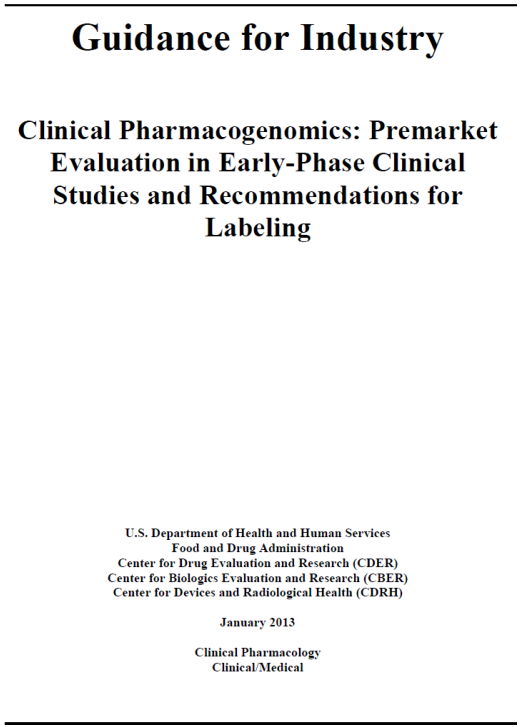
Figure: Simulation Study Results (Adjusted Rand Index) for Scenarios 1, 2, 3

Simulation Results

Scen.	α	t_{full} L_2	t_{full} CORT	t_{full} COR	t_{full} DTW	**pkDip**	t_{sparse} L_2	t_{sparse} CORT	t_{sparse} COR	t_{sparse} DTW
3′	0.01	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	0.05	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.98 (0.10)	**1.00 (0.01)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)
	0.10	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.41 (0.37)	**0.92 (0.06)**	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.50 (0.44)
	0.15	0.95 (0.13)	0.97 (0.05)	0.94 (0.06)	0.00 (0.11)	**0.83 (0.12)**	0.79 (0.36)	0.95 (0.08)	0.94 (0.09)	0.06 (0.20)
	0.20	0.67 (0.37)	0.83 (0.18)	0.79 (0.17)	0.00 (0.06)	**0.76 (0.13)**	0.39 (0.42)	0.72 (0.32)	0.78 (0.20)	0.00 (0.08)
	0.25	0.34 (0.35)	0.59 (0.29)	0.62 (0.21)	0.00 (0.05)	**0.61 (0.25)**	0.16 (0.27)	0.40 (0.36)	0.59 (0.24)	0.00 (0.06)
	0.30	0.18 (0.25)	0.42 (0.29)	0.50 (0.21)	0.00 (0.04)	**0.50 (0.23)**	0.06 (0.15)	0.29 (0.30)	0.48 (0.21)	0.00 (0.05)
	0.50	0.04 (0.09)	0.13 (0.16)	0.25 (0.16)	0.00 (0.05)	**0.21 (0.14)**	0.02 (0.07)	0.05 (0.11)	0.27 (0.14)	0.00 (0.05)
4	0.01	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	**1.00 (0.00)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
	0.05	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.91 (0.16)	**0.95 (0.12)**	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)
	0.10	0.90 (0.13)	0.93 (0.10)	0.92 (0.05)	0.36 (0.18)	**0.53 (0.06)**	0.70 (0.19)	0.77 (0.20)	0.91 (0.05)	0.64 (0.18)
	0.15	0.60 (0.10)	0.64 (0.12)	0.71 (0.10)	0.08 (0.10)	**0.47 (0.04)**	0.54 (0.05)	0.56 (0.05)	0.70 (0.09)	0.28 (0.18)
	0.20	0.49 (0.09)	0.51 (0.05)	0.54 (0.07)	0.02 (0.03)	**0.43 (0.06)**	0.46 (0.13)	0.49 (0.05)	0.54 (0.07)	0.06 (0.09)
	0.25	0.37 (0.17)	0.44 (0.05)	0.45 (0.06)	0.01 (0.02)	**0.40 (0.04)**	0.24 (0.21)	0.43 (0.07)	0.45 (0.06)	0.03 (0.04)
	0.30	0.18 (0.19)	0.38 (0.08)	0.37 (0.05)	0.01 (0.01)	**0.35 (0.07)**	0.07 (0.15)	0.33 (0.13)	0.37 (0.05)	0.01 (0.02)
	0.50	0.00 (0.01)	0.10 (0.11)	0.18 (0.06)	0.00 (0.00)	**0.05 (0.06)**	0.00 (0.00)	0.06 (0.09)	0.19 (0.05)	0.00 (0.01)

Figure: Simulation Study Results (Adjusted Rand Index) for Scenarios 3’ and 4

Genetic Information and PK



Guidance for Industry

Clinical Pharmacogenomics: Premarket Evaluation in Early-Phase Clinical Studies and Recommendations for Labeling

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

January 2013

Clinical Pharmacology
Clinical/Medical

10300.fnl.doc

… In general, the effects of genetic differences on PK are easier to characterize because they affect a readily measured feature of the drug (its pharmacokinetics). … In many cases, the mechanism that causes differences in PK is related to metabolizing enzymes or transport proteins with well-established genetic polymorphisms, as is the case for CYP2C9, CYP2C19, CYP2D6, or SLCO1B1, so that such differences can be anticipated.

Using Genetic Information to Guide Drug Therapy

Pharmacogenetic tests, along with other information about patients and their disease or condition, can play an important role in drug therapy. When a health care provider is considering prescribing a drug, knowledge of a patient's genotype may be used to aid in determining a therapeutic strategy, determining an appropriate dosage, or assessing the likelihood of benefit or toxicity.

For specific pharmacogenetic associations, the FDA has evaluated and believes there is sufficient scientific evidence to suggest that subgroups of patients with certain genetic variants, or genetic variant-inferred phenotypes (such as affected subgroup in the table of next slide), are likely to have altered drug metabolism, and in certain cases, differential therapeutic effects, including differences in risks of adverse events.

https://www.fda.gov/medical-devices/precision-medicine/table-pharmacogenetic-associations