Optimal designs for dose-response models with partially observed interim/hidden layers

Valerii V. Fedorov

ICON plc, North Wales, PA, USA

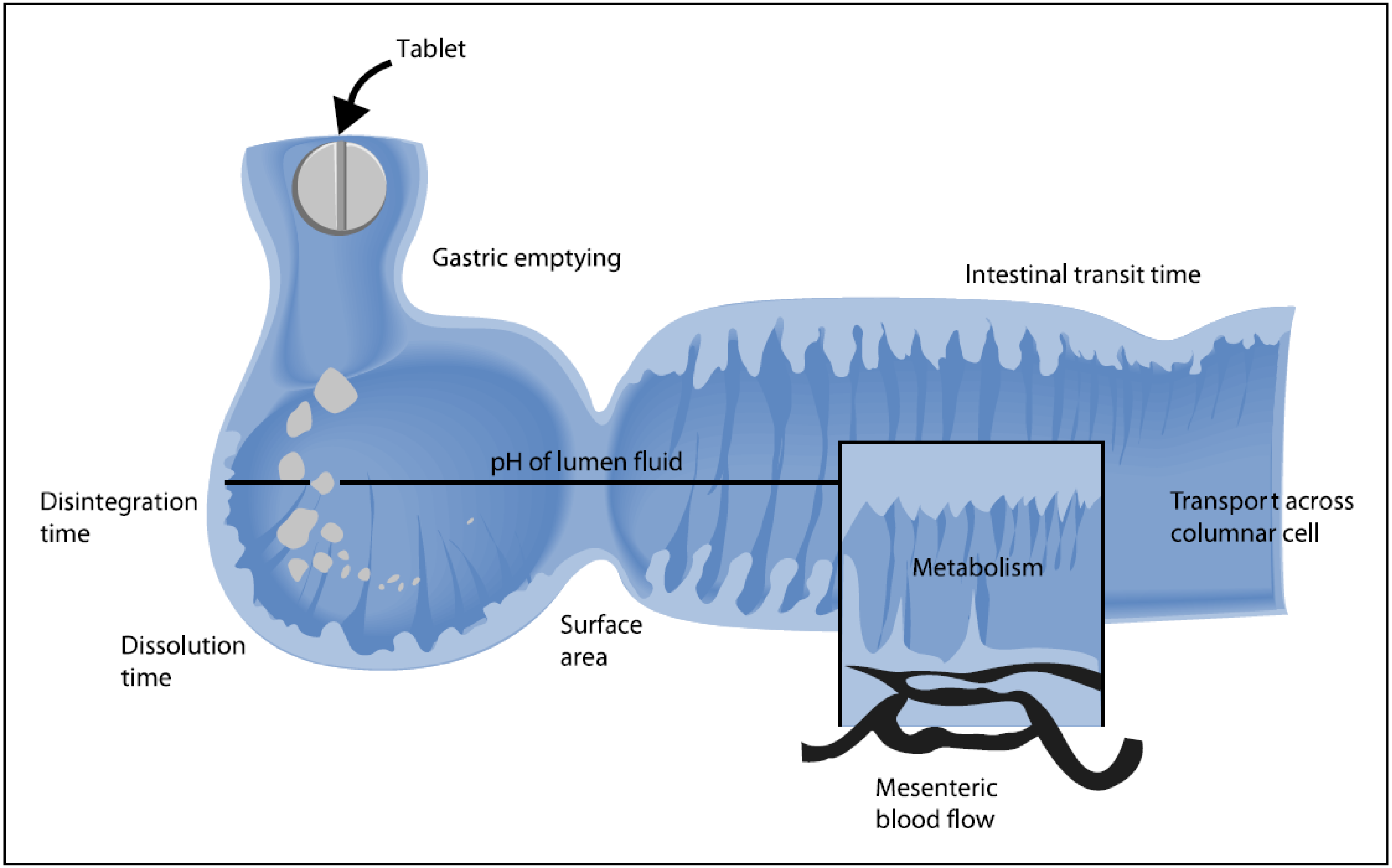
Abstract

In dose selection/ranging trials a researcher knows the dose given, may/may not measure (PK stage or layer) certain characteristics, such as AUC, Cmax, or Tmax, and observes the response(s) to treatment, for example, efficacy and /or toxicity end-points (PD stage or layer). For every stage its own model can be suggested and outputs from the first one can be viewed as candidate inputs for the second stage model. In cases when outcomes of the first stage cannot be observed the setting reminds the neural networks modeling and that partially explains our terminology. In this presentation various designs will be compared and discussed.

Acknowledgement

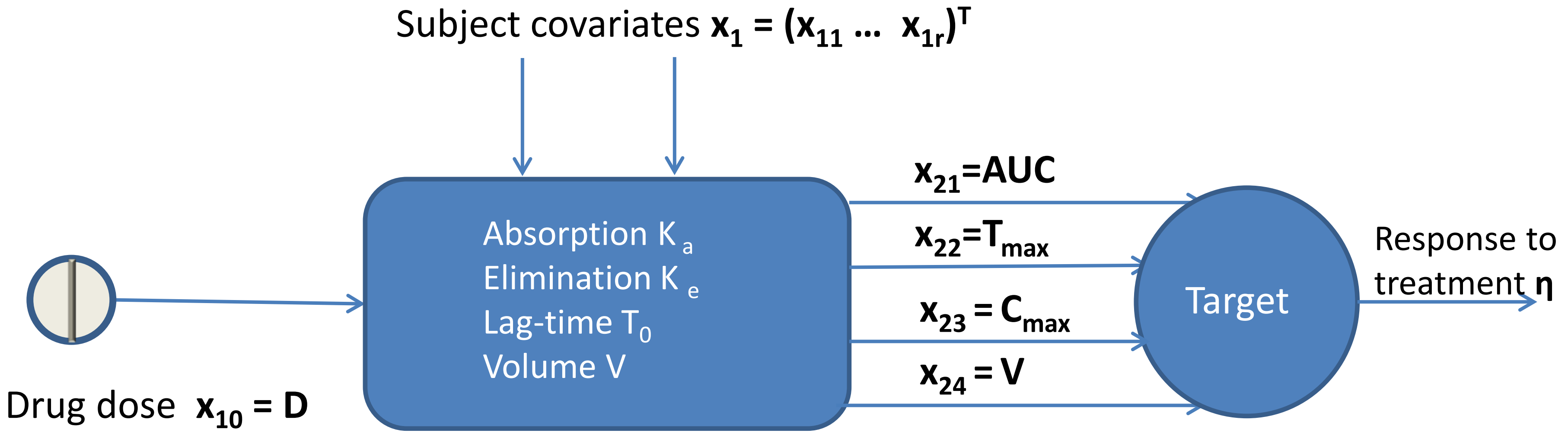
My sincere appreciation to Prof. Anastasia Ivanova for her efforts in starting this projects

Pharmaceutical Press www.pharmpress.com: *"Basic Pharmacokinetics"*



Passage of drug in the gastrointestinal tract until transport across the membrane

Simple model of the previous and problem setting



Subject covariates $\mathbf{x_1 = (x_{11} \ldots\ x_{1r})^T}$

Absorption K_a
Elimination K_e
Lag-time T_0
Volume V

$\mathbf{x_{21}=AUC}$

$\mathbf{x_{22}=T_{max}}$

$\mathbf{x_{23} = C_{max}}$

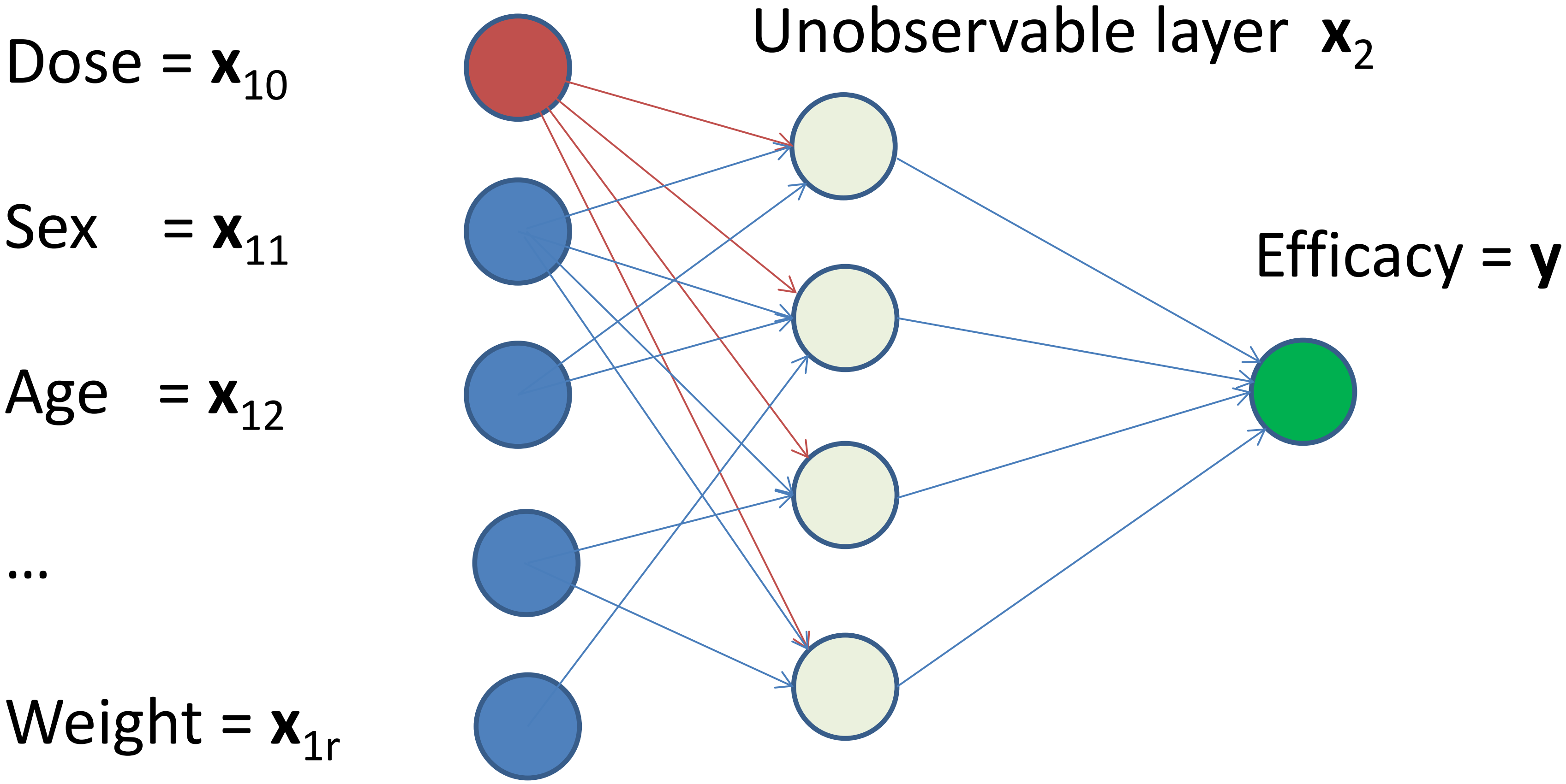
$\mathbf{x_{24} = V}$

Target

Response to treatment $\boldsymbol{\eta}$

Drug dose $\mathbf{x_{10} = D}$

Machine learning setting



Dose = $\mathbf{x}_{10}$

Sex = $\mathbf{x}_{11}$

Age = $\mathbf{x}_{12}$

…

Weight = $\mathbf{x}_{1r}$

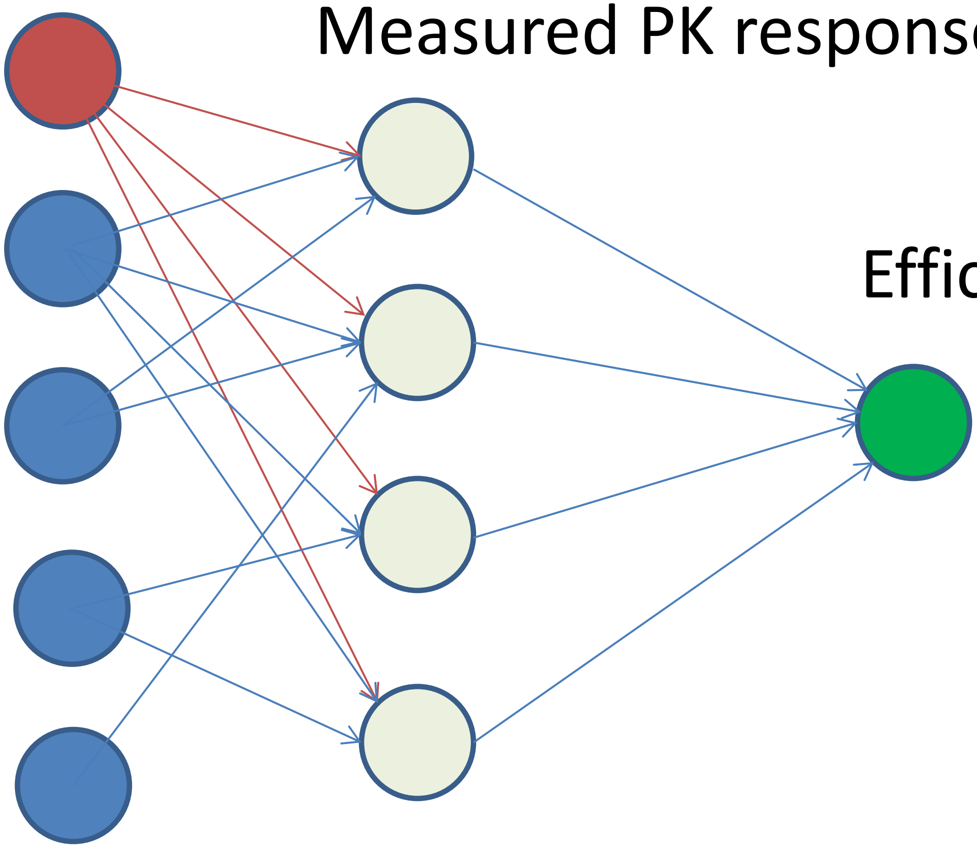
Unobservable layer $\mathbf{x}_2$

Efficacy = $\mathbf{y}$

$\mathbf{S}_{21}\,\mathbf{x}_1 = \mathbf{x}_2$

$\mathbf{S}_{32}\,\mathbf{x}_2 = \mathbf{y}$

Enhanced learning with models



Dose = $\mathbf{x}_{10}$

Sex = $\mathbf{x}_{11}$

Age = $\mathbf{x}_{12}$

…

Weight = $\mathbf{x}_{1r}$

Measured PK response $\mathbf{x}_2$

Efficacy = $\mathbf{y}$

$\mathbf{S}_{21}(\beta,\mathbf{x}_1) = \mathbf{x}_2$ $\mathbf{S}_{32}(\gamma,\mathbf{x}_2) = \mathbf{y}$

PK measurements: one compartment model

Building $\mathbf{S}_{21}$:

$$\begin{cases} \dot{d}_0(t) = & -K_a d_0(t), & & d_0(0) = D \\ \dot{d}_1(t) = & K_a d_0(t) & - K_e d_1(t), & d_1(0) = 0 \end{cases}$$

$K_a,\ K_e,\ V$ are functions of $\underline{\mathbf{x}_1^T} = (x_{11}, \ldots, x_{1r})$

$$\boxed{\begin{aligned} AUC(T) &= \int_0^T C(t,\gamma)dt = \frac{DK_a}{V(K_aT - K_eT)}\left[\frac{1-e^{-K_e}}{K_e} - \frac{1-e^{-K_a}}{K_a}\right] \\ AUC &= AUC(\infty) = \frac{D}{VK_e} \\ C_{max} &= \max_t C(t,\gamma) = \frac{D}{V}\left(\frac{K_a}{K_e}\right)^{-K_e/(K_a-K_e)} \\ T_{max} &= \arg\max_t C(t,\gamma) = \frac{lnK_a - \ln K_e}{K_a - K_e} \end{aligned}}$$

$$\boxed{\mathbf{x}_2^T = (1,\ AUC,\ C_{max},\ T_{max})}$$

PK measurements: simpler approach

$$\mathbf{x}_2^T = (1,\ AUC,\ C_{max},\ T_{max})$$

$$\Big\downarrow$$

$$\begin{aligned} AUC &= D\varphi_1(\beta_1, \underline{\mathbf{x}}_1)/V \\ C_{max} &= D\varphi_2(\beta_2, \underline{\mathbf{x}}_1)/V \\ T_{max} &= \varphi_3(\beta_3, \underline{\mathbf{x}}_1) \\ V &= \varphi_4(\beta_4, \underline{\mathbf{x}}_1) \end{aligned}$$

- *We use some information from the previous slide*
- *Intrinsic relations between AUC, C_{max} and T_{max} are not taken account*
- *AUC, C_{max} and T_{max} are random*

PD stage: logit model

Building $\mathbf{S}_{32}$ for binary end-point:

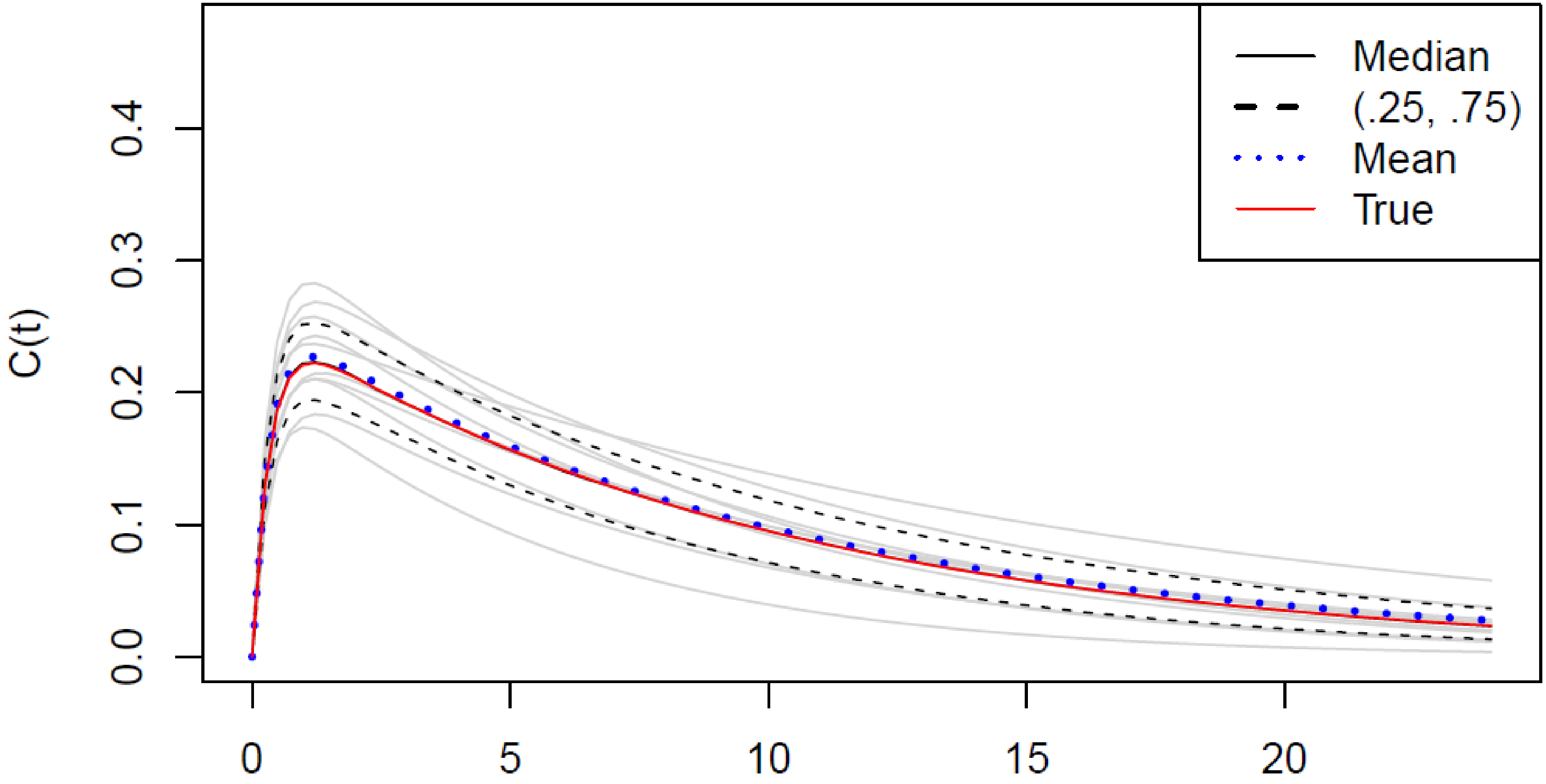
$$\begin{aligned} \mathrm{Prob}(Y=1) &= p(\mathbf{x}_2, \boldsymbol{\theta}) \\ \mathrm{Prob}(Y=0) &= 1 - p(\mathbf{x}_2, \boldsymbol{\theta}) \end{aligned}$$

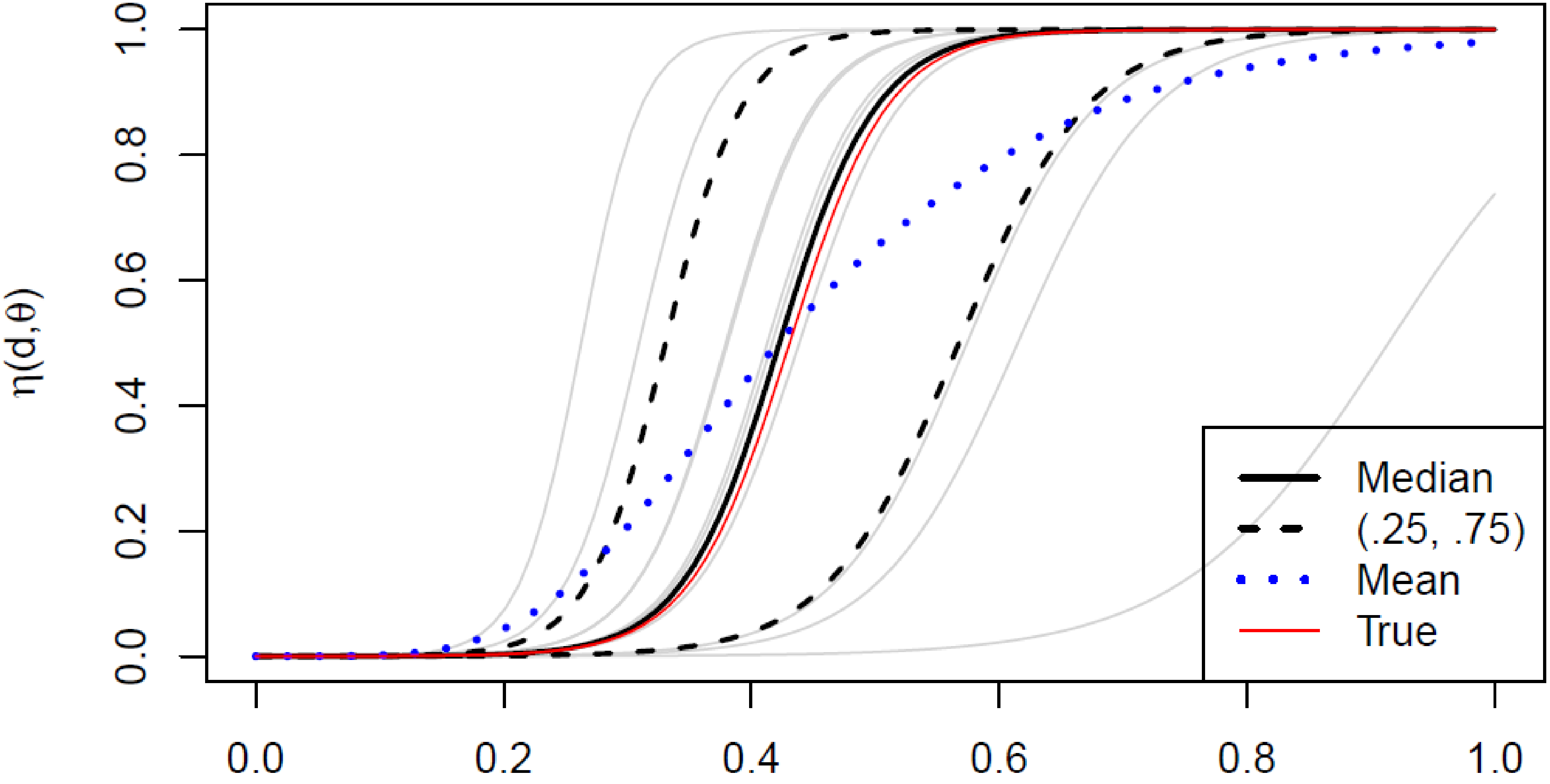
$$p(\mathbf{x}_2, \boldsymbol{\theta}) = \frac{\exp \boldsymbol{\theta}^T \mathbf{f}(\mathbf{x}_2)}{1 + \exp \boldsymbol{\theta}^T \mathbf{f}(\mathbf{x}_2)}$$

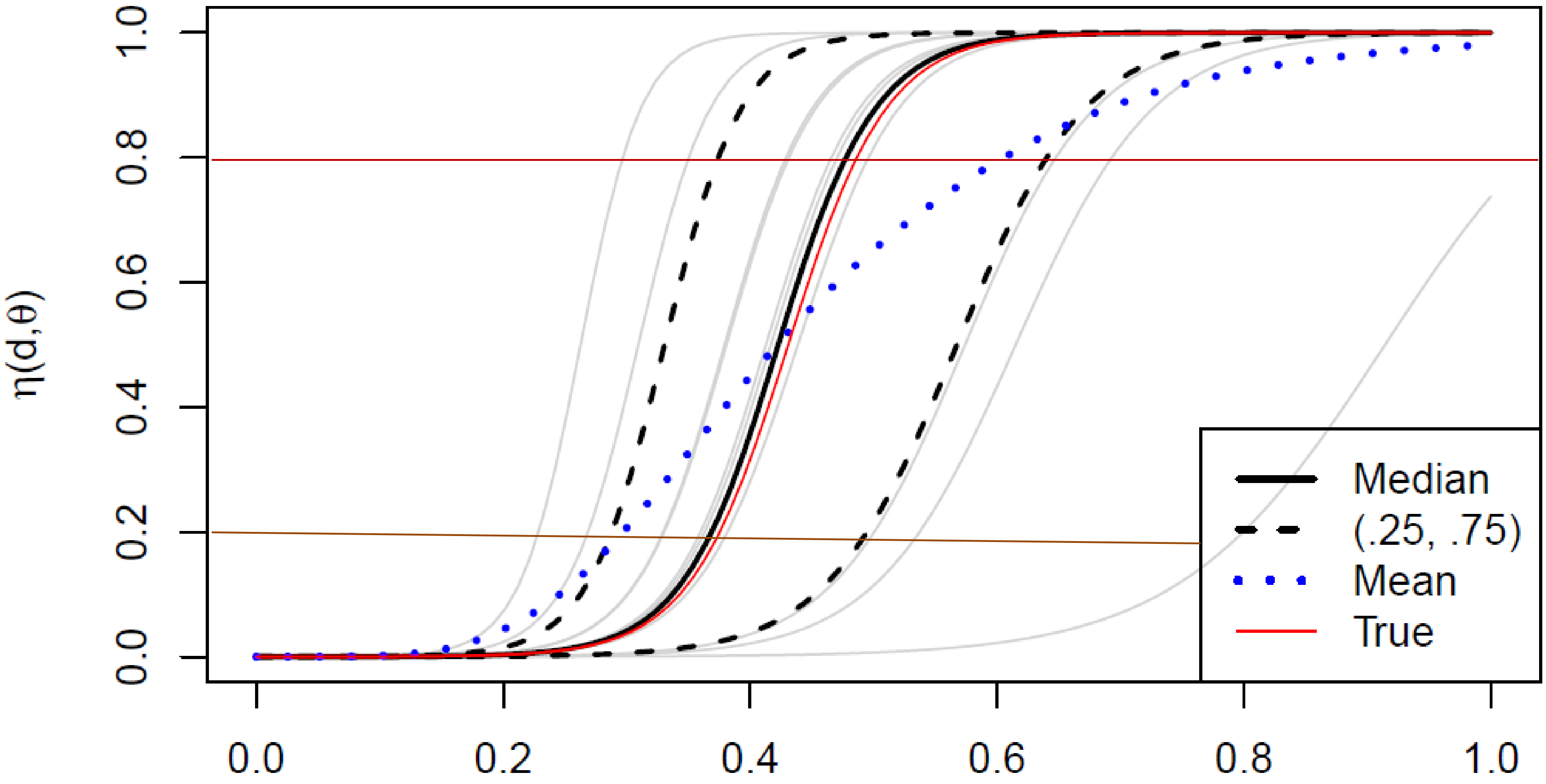
$$\begin{aligned} \boldsymbol{\theta}^T \mathbf{f}(\mathbf{x}_2) &= \theta_1 + \theta_2 AUC + \theta_3 C_{max} + \theta_4 T_{max} \\ &= \theta_1 + \theta_2 D \times \underline{AUC} + \theta_3 D \times \underline{C_{max}} + \theta_4 T_{max} \end{aligned}$$

Can be controlled

Random


C(t)
0.0
0.1
0.2
0.3
0.4
0
5
10
15
20
Median
(.25, .75)
Mean
True


η(d,θ)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Median
(.25, .75)
Mean
True


η(d,θ)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Median
(.25, .75)
Mean
True

I. Trial design focusing on ϴ

Fisher Information Matrix (FIM) for a single observation:

$$\begin{aligned}\boldsymbol{\mu}(\mathrm{x}_2, \boldsymbol{\theta}) &= \frac{1}{p(\mathrm{x}_2, \boldsymbol{\theta})(1 - p(\mathrm{x}_2, \boldsymbol{\theta}))} \frac{\partial p(\mathrm{x}_2, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial p(\mathrm{x}_2, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \\ &= p(\mathrm{x}_2, \boldsymbol{\theta})(1 - p(\mathrm{x}_2, \boldsymbol{\theta}))\mathbf{f}(\mathrm{x}_2)\mathbf{f}^T(\mathrm{x}_2)\end{aligned}$$

Total FIM :

$$\begin{aligned}\mathcal{M}(\xi_N, \boldsymbol{\theta}) &= \sum_{i=1}^{r} \sum_{j=1}^{n_i} \boldsymbol{\mu}(\mathrm{x}_{2,ij} \boldsymbol{\theta}) \\ &= N \sum_{i=1}^{r} \sum_{j=1}^{p_i} \boldsymbol{\mu}(\mathrm{x}_{2,ij} \boldsymbol{\theta}) = NM(\xi_N, \boldsymbol{\theta})\end{aligned}$$

By the strong law of large numbers:

$$\boxed{M(\xi_N, \boldsymbol{\theta}) \Longrightarrow M(\xi, \boldsymbol{\theta}) = \sum_{i}^{r} \pi_i \boldsymbol{\mu}(D_i, \boldsymbol{\theta})}$$

$$\boldsymbol{\mu}(D_i, \boldsymbol{\theta}) = \mathrm{E}\left[\boldsymbol{\mu}(\mathrm{x}_{2,ij} \boldsymbol{\theta})\right] \qquad \mathrm{x}_2^T = (1,\ AUC,\ C_{max},\ T_{max})$$

Atkinson A. and Fedorov V. (1988), The Optimum Design of Experiments in the Presence of Uncontrolled Variability and Prior Information, North-Holland, pp. 327-344

Atkinson A., Fedorov V., Herzberg A., Zhang R. (2014), Elemental information matrices and optimal experimental design for generalized regression models, *JSPI*, **144**: *81–91.*

II. Trial design focusing on Θ

Main optimization problem:

$$\boxed{\xi^* = \arg\min_{\xi} \Psi\left[M(\xi, \boldsymbol{\theta})\right]}$$

- All standard results of convex design theory are valid, for instance, the equivalence theorem (D-criterion):

$$\mathrm{tr}\left[\boldsymbol{\mu}(D, \boldsymbol{\theta}) M^{-1}(\xi^*, \boldsymbol{\theta})\right] \leq \dim \boldsymbol{\theta}, \quad D \in \mathcal{D}$$

- Computing of

$$\boldsymbol{\mu}(D, \boldsymbol{\theta}) = \mathrm{E}\left[\boldsymbol{\mu}(\mathrm{x}_2, \boldsymbol{\theta})\right]$$

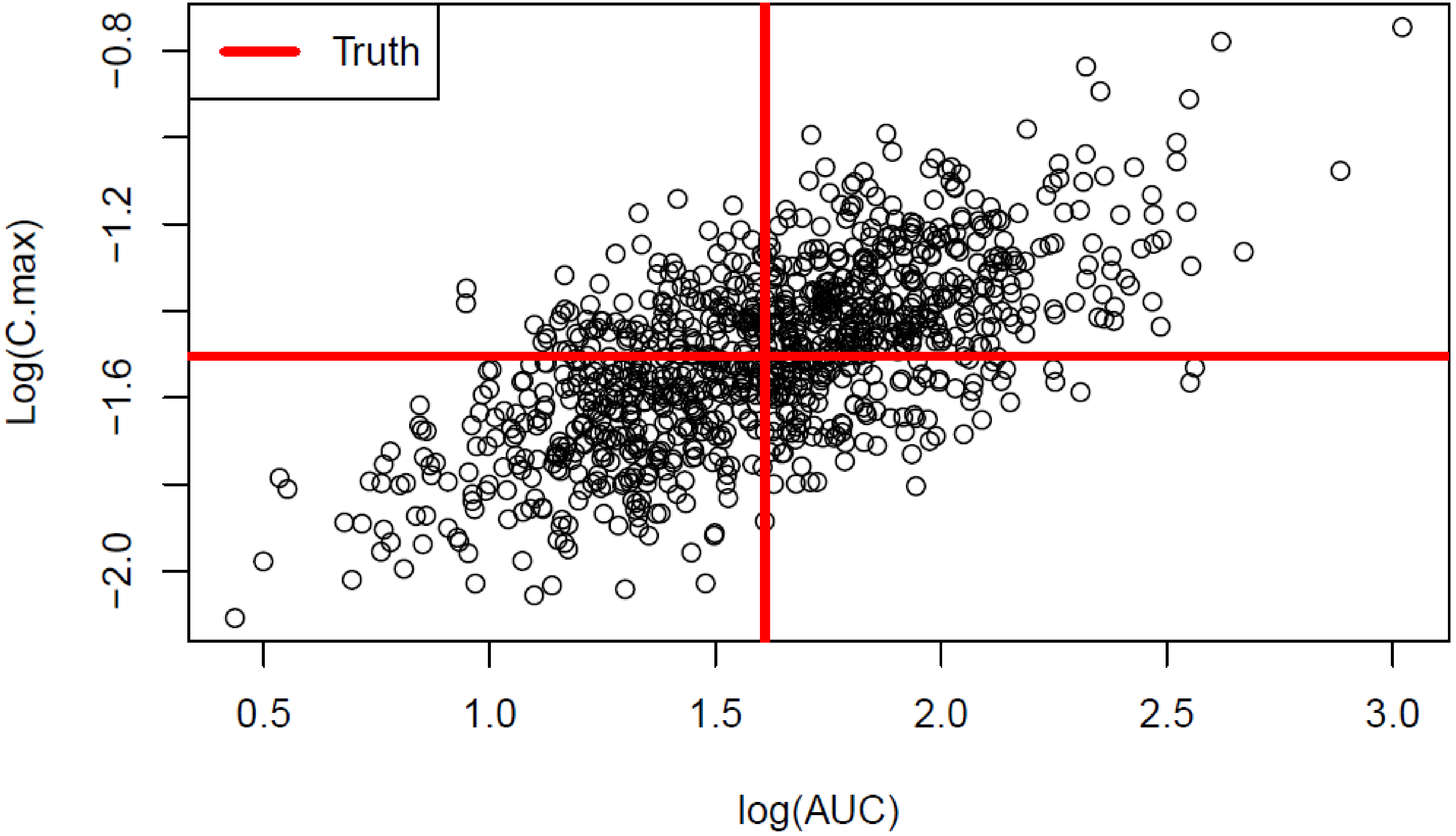
  is the major challenge.
- All inputs uncontrolled inputs **x**$_1$ and **x**$_2$ are available on the trial completion and the standard (conditional) likelihood method for multivariate logit model is applicable
- The approach removes most of "between subjects" variability and that increases the predictive power.

Next steps

- Fusing both stages in single design/estimation procedure
- Combining the enhanced network training with the standard one (no interim observation)
- Comparison of the proposed approach with approaches based on models that appear in the Poisson sampling scheme*

*Johnson N., Kotz S. and Kemp A. (1993). Univariate Discrete Distributions, Ch. 12.2, Second Edition, Wiley

Back up



Moment based design

$$\mathrm{E}\left[Y_{ij}\right] = \bar{p}(D_i, \boldsymbol{\theta}) = \mathrm{E}\left[p(\mathbf{x}_{ij}, \boldsymbol{\theta})\right]$$

$$\boldsymbol{\Sigma}_i = \mathrm{Var}\left[Y_{ij}\right] = \bar{p}(D_i, \boldsymbol{\theta})(1 - \bar{p}(D_i, \boldsymbol{\theta}))$$

$$\boldsymbol{\mu}(D_i, \boldsymbol{\theta}) = \frac{\partial \bar{p}(D_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \bar{p}(D_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$$

Standard design problem

Iterated least square method

It is not the Fisher information matrix ! See Zegers P. (2015), Entropy, **17**, 4918-4939