Mixed-Effects Model Qualification Using Predictive Check

Pravin R Jadhav

Graduate Student,

Virginia Commonwealth University, Richmond, VA

Office of Clinical Pharmacology and Biopharmaceutics, CDER / FDA

Outline

- Posterior Predictive Check (PPC)
- Predictive check theory
- Simulation example
- Applications
- Inference

(Posterior) Predictive Check (PPC)

- A method to check if the posited model should be excluded because it fails to provide a reasonable summary of the data at hand
- A summary feature of the data (aka statistic) is calculated to check the compatibility of model and data
- The present work does not consider the posterior distribution of model parameters

A summary feature (a statistic)

- Test statistic- T(y)

  Does not depend on model parameters

  - Average concentration at specific time (Cmax, Cmin, Css, Ct)
  - Area under the curve (AUC)

- Discrepancy variable- T(y,θ)

  Depends on model parameters

  - Sum of squared errors $SSE = \sum(\hat{C}_t - C_t)^2$

Predictive check

- Posterior distribution accounts for the parameter uncertainty and prior knowledge
  - Not considered in the current effort
- Maximum Likelihood (ML) estimation ignores uncertainty and prior
  - ML widely used in population PK/PD data analysis
- So, not exactly- Posterior
- More appropriately,

ML **Predictive check**

PPC- Yano, Beal, and Sheiner 2001

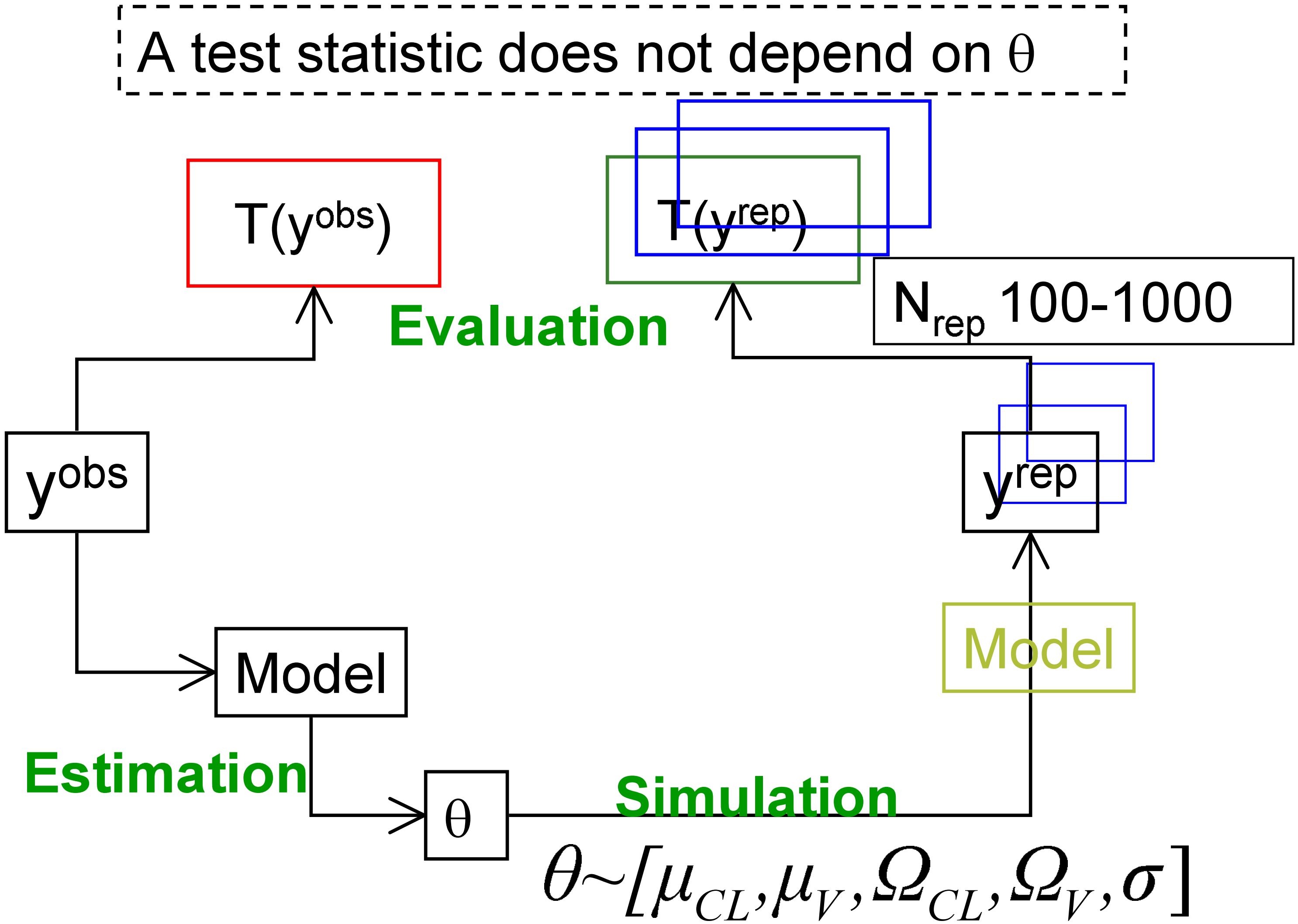
- First attempt to systematically evaluate properties of PPC for PK/PD model selection
- Results applicable to single subject setting and application to population data analysis with several covariates is not very obvious
- No clear distinction between the properties of a test statistic versus a discrepancy variable
- Only the p-value (Pp) was used for predictive performance- other alternatives were not explored

Inferences

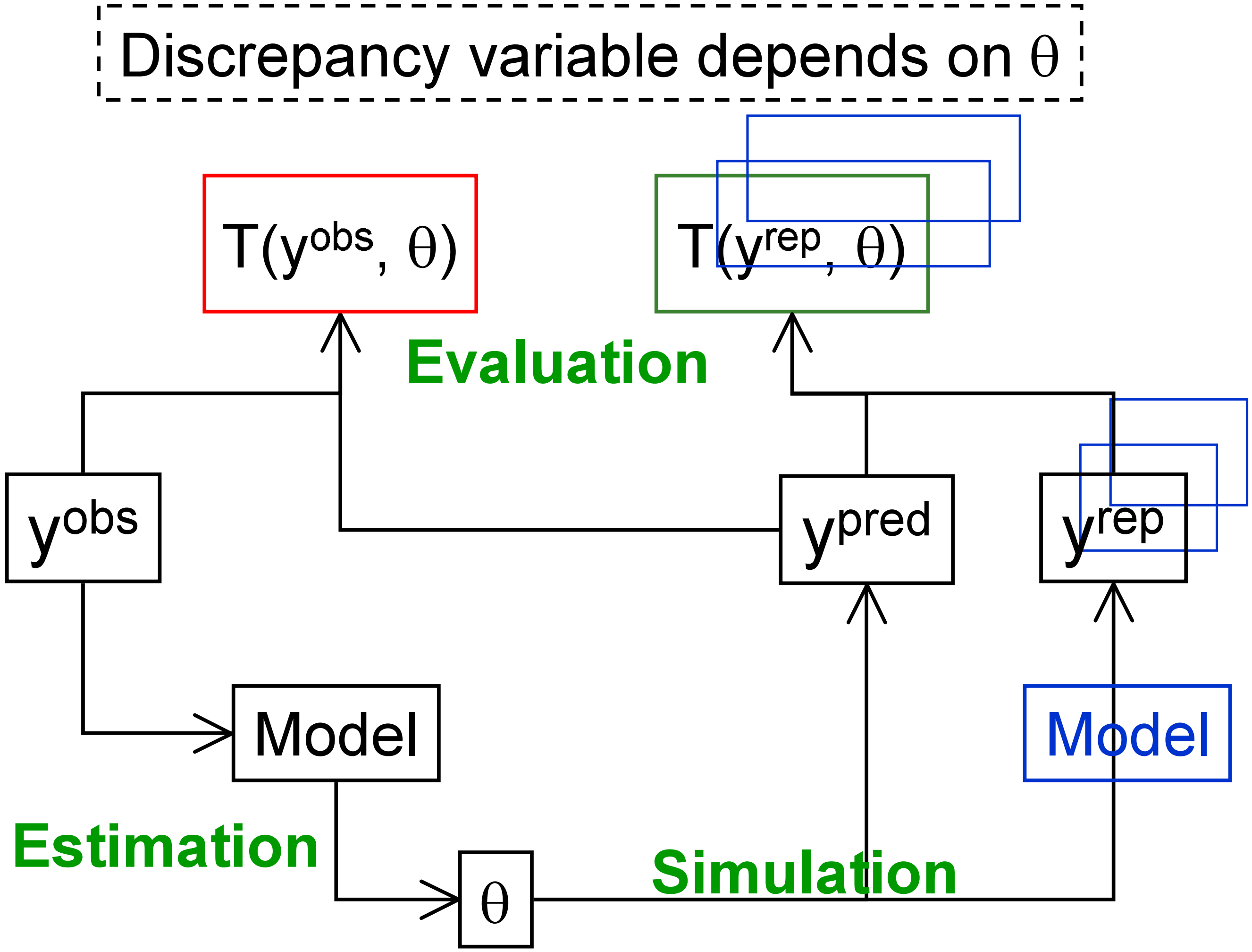
- Use of discrepancy variable cannot aid in rejecting false models
- Use of test statistic can aid in rejecting false models
  - Selection of informative test statistic is challenging
- 95% prediction interval cannot aid in rejecting false models
- Equivalence based comparison of test statistic is more informative than significance based comparison

Predictive Check Theory

Theory- three steps in qualification

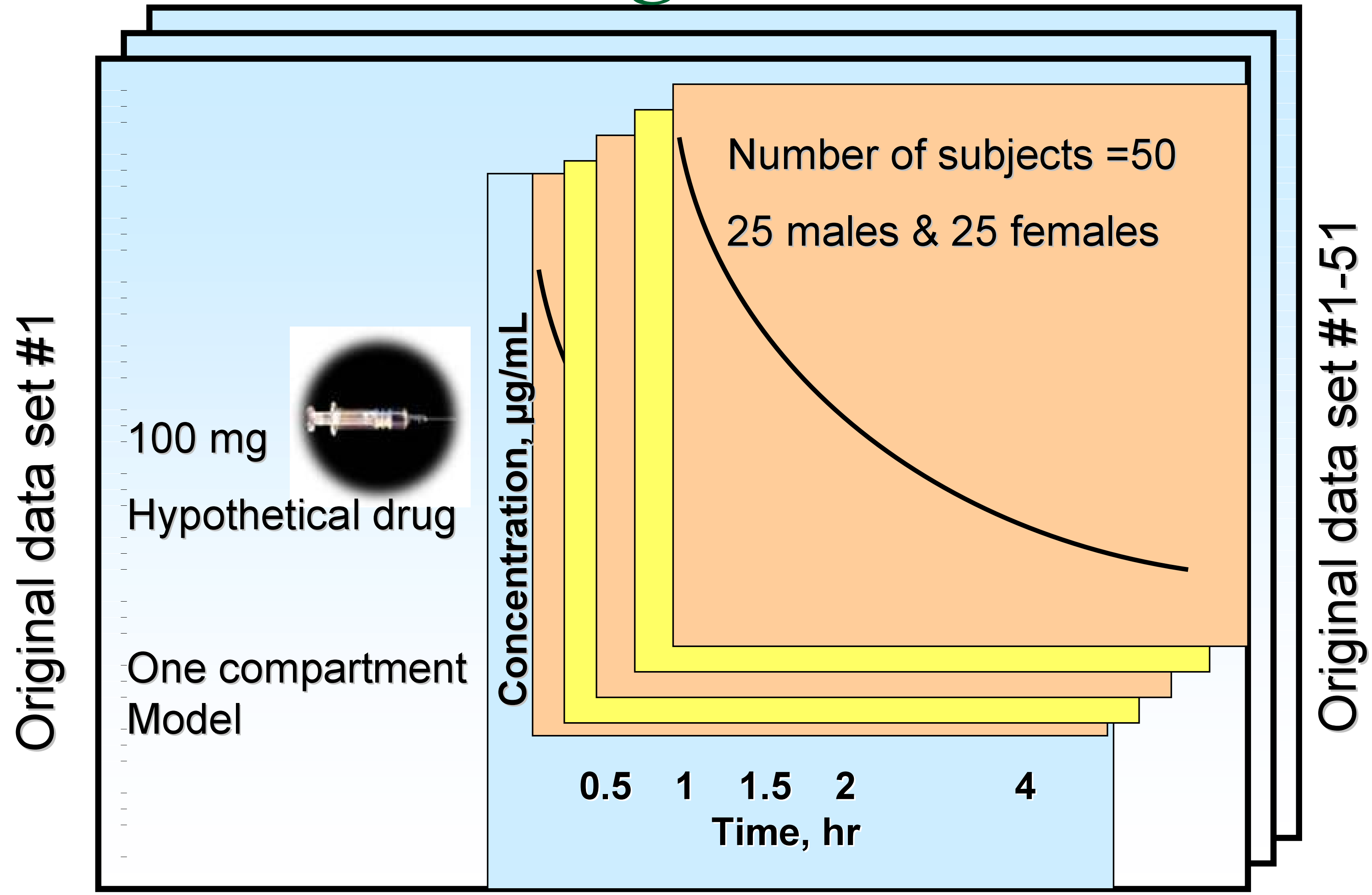


Evaluation cont'd- Discrepancy variable



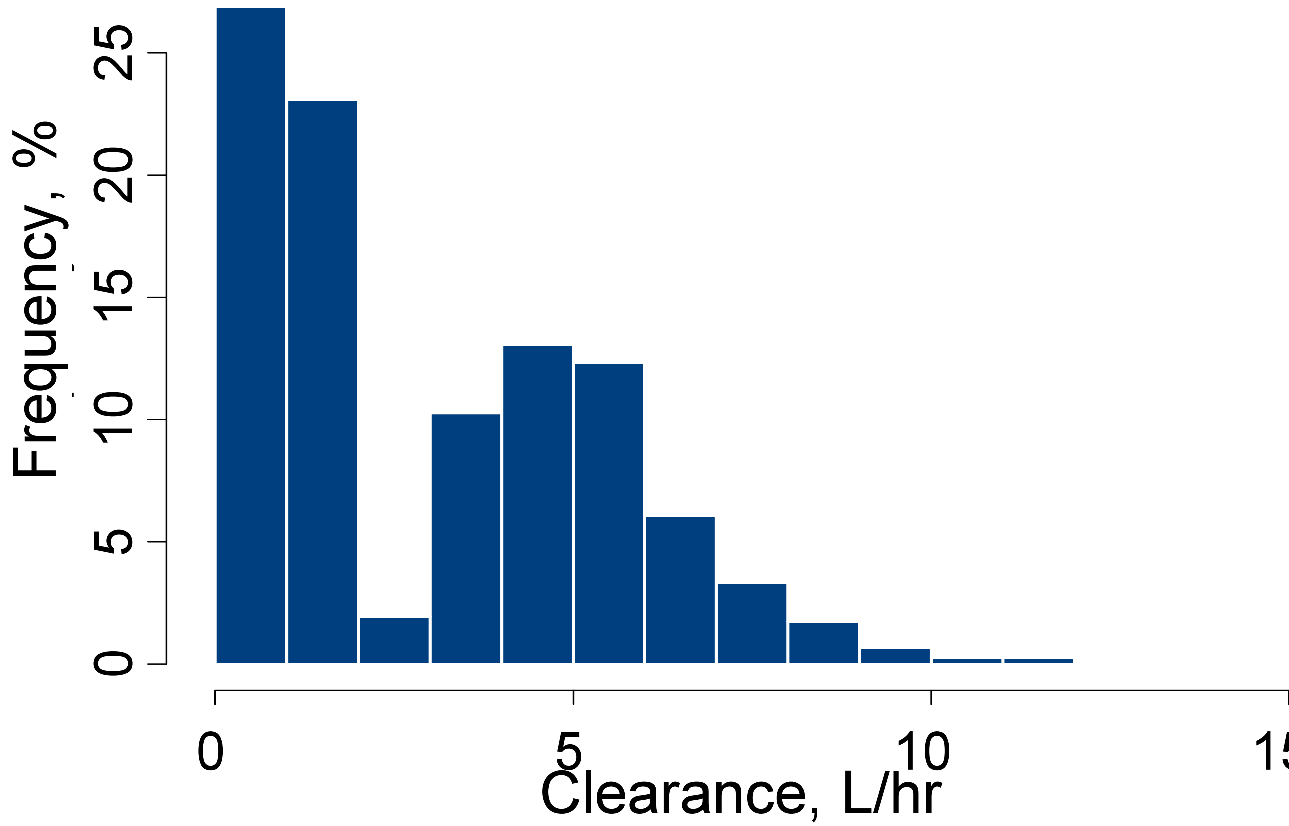
Simulation example

Simulation setting



Simulation setting

- Input model- One compartment PK model
- CL(males)=5·CL(females)



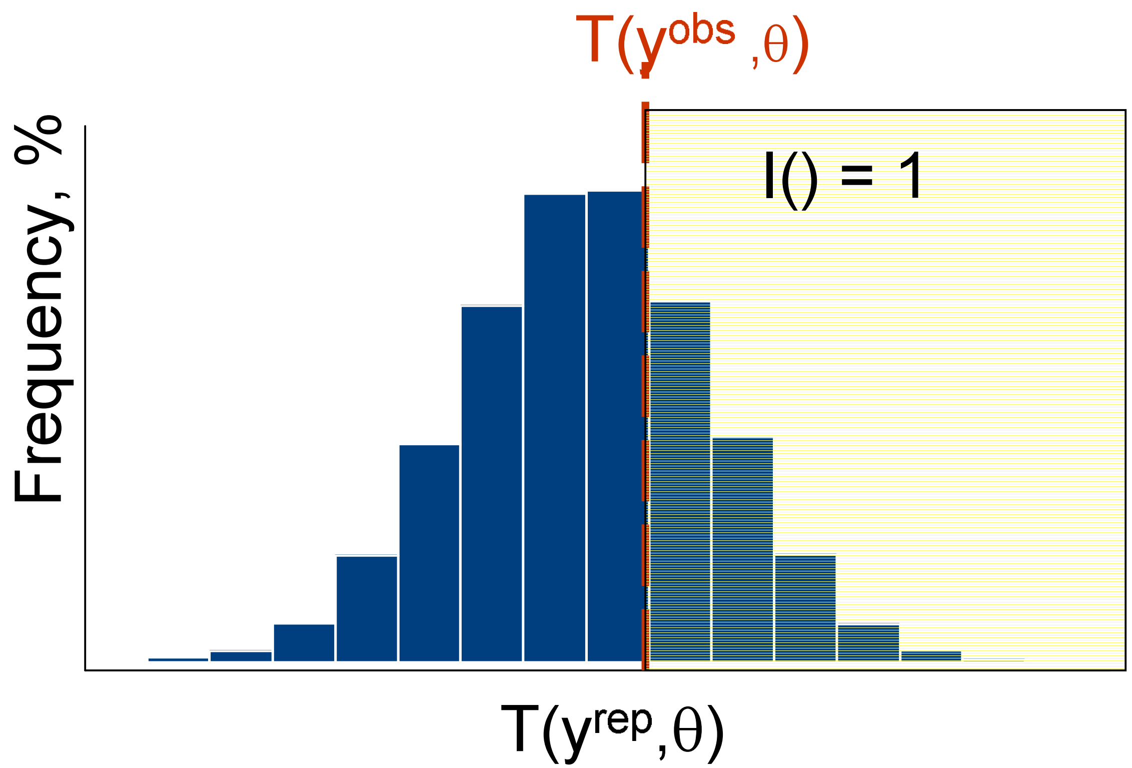
Predictive check for one dataset

- Estimation step: Obtain estimates of θ
  - **“False model” (males=females)**
  - **“True model” (males $\neq$ females)**
- Simulation step: 200 simulations for each model
- Evaluation step:
  - **Test statistic**
    1. **C (t=0.5 hr)**
    2. **C (t=4 hr)**
  - **Discrepancy variable**
    1. **Sum of squared errors (SSE)**

Evaluation- selection criteria

- Predictive p-value (Pp)
- Probability of equivalence (p^{eqv})

Predictive p-value (Pp)

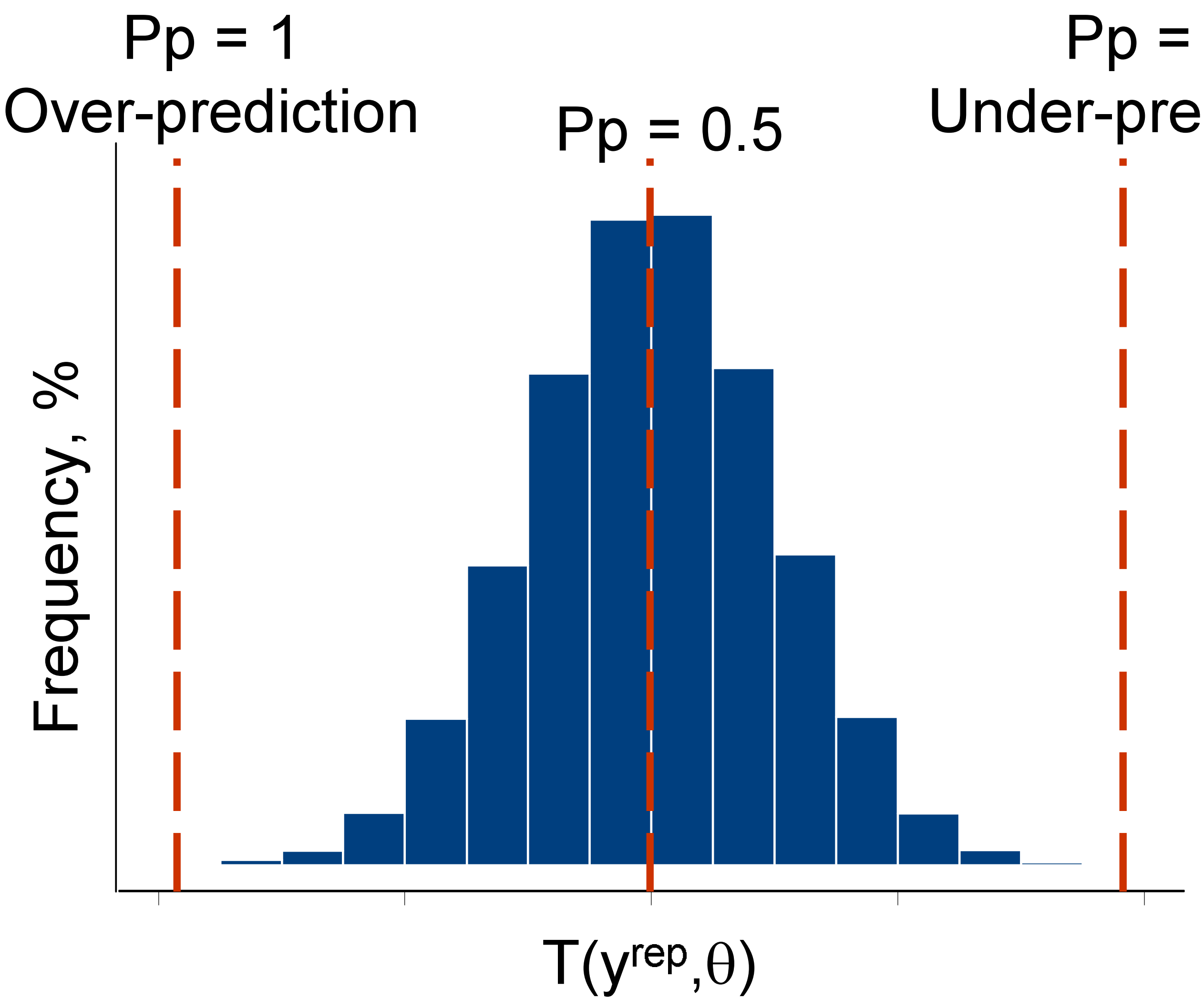


Rep	I()
1	0
2	1
…	…
200	1

$\overline{I()}$ **=Pp**

$$Pp = \frac{1}{N_{rep}} \sum_{i=1}^{N_{rep}} I(T(y_i^{rep}, \theta) \geq T(y^{obs}, \theta))$$

What is the ideal value for Pp?



Optimal Pp is subjective.

Can be driven by clinical relevance.

T(y,θ) is uninformative

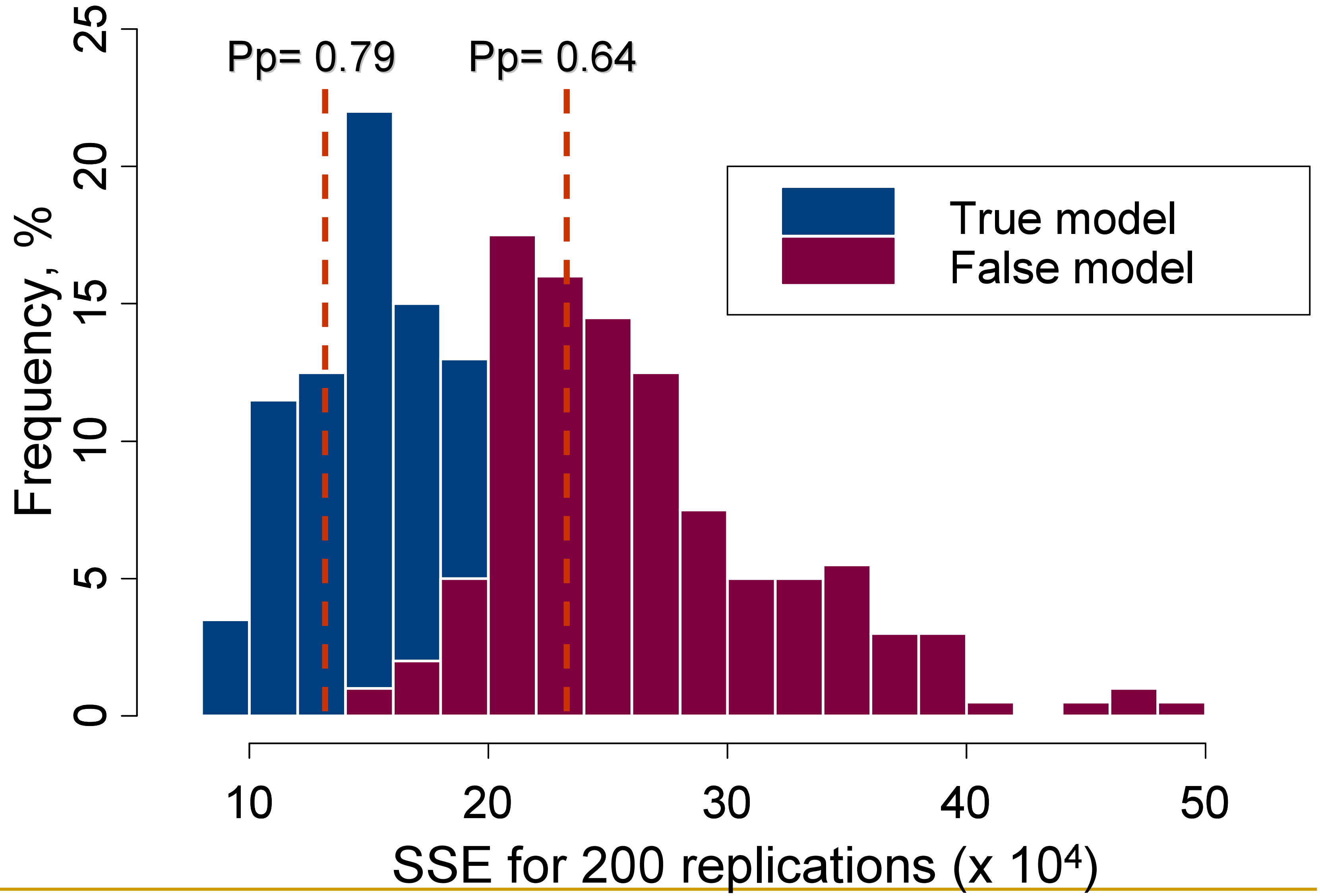
- Sum of squared error (SSE), T(y,θ), is used to calculate Pp
- Discrepancy variable is scaled relative to y^{pred}
  - Caution should be applied while comparing models

$$SSE_{False}^{obs} = 23.3 \times 10^4 \qquad SSE_{False}^{rep} = 24.4 \times 10^4$$

$$SSE_{True}^{obs} = 13.4 \times 10^4 \qquad SSE_{True}^{rep} = 10.9 \times 10^4$$

$$Ratio^{obs} \sim 2 \qquad Ratio^{rep} \sim 2$$

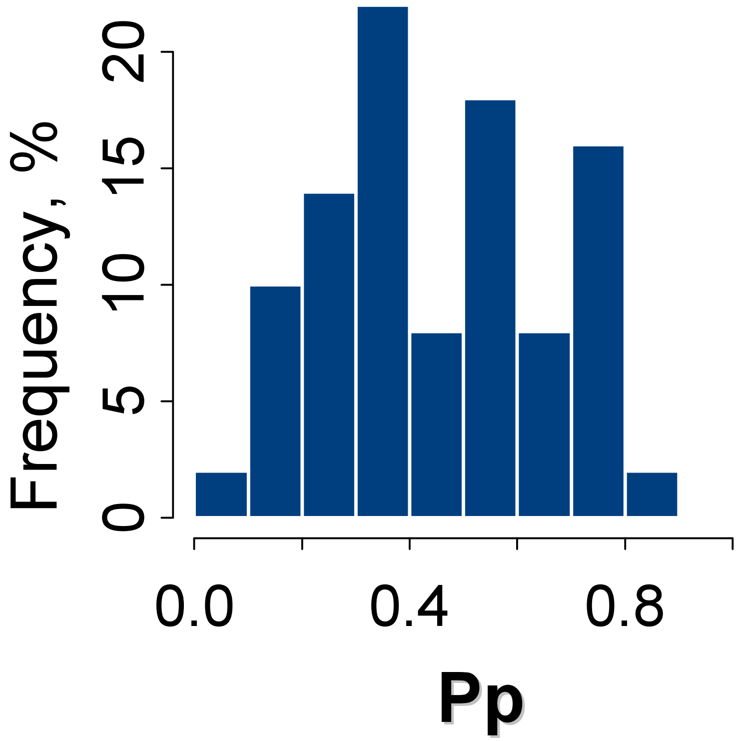
T(y,θ) is uninformative



Predictive p-value (Pp)

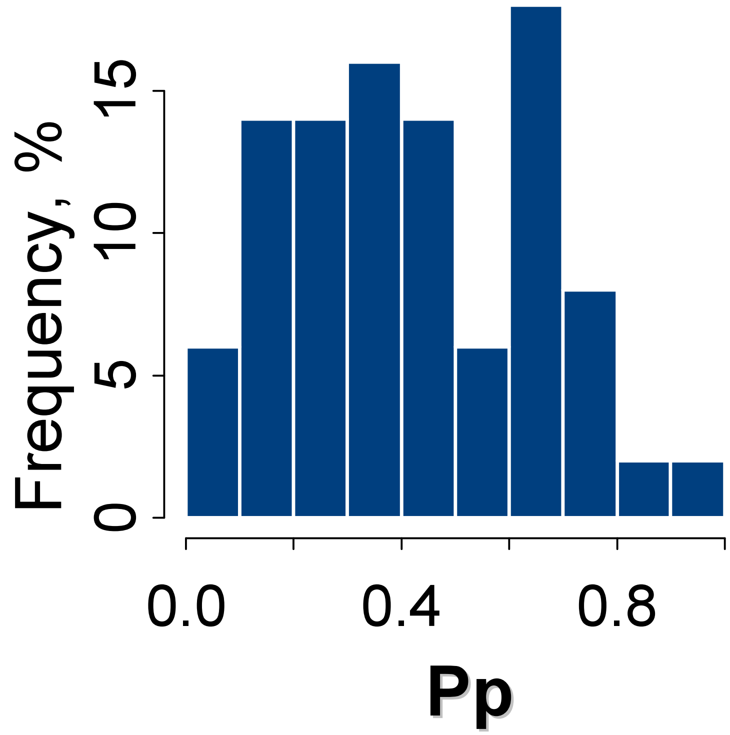
- Use of discrepancy variable cannot help in rejecting model.
- Experiment is repeated for 51 original datasets .
  - 200 replications for each original dataset
- Thus, we will have 51 Pp value
- What would be the shape of Pp distribution?

Predictive p-value (Pp)



? Model

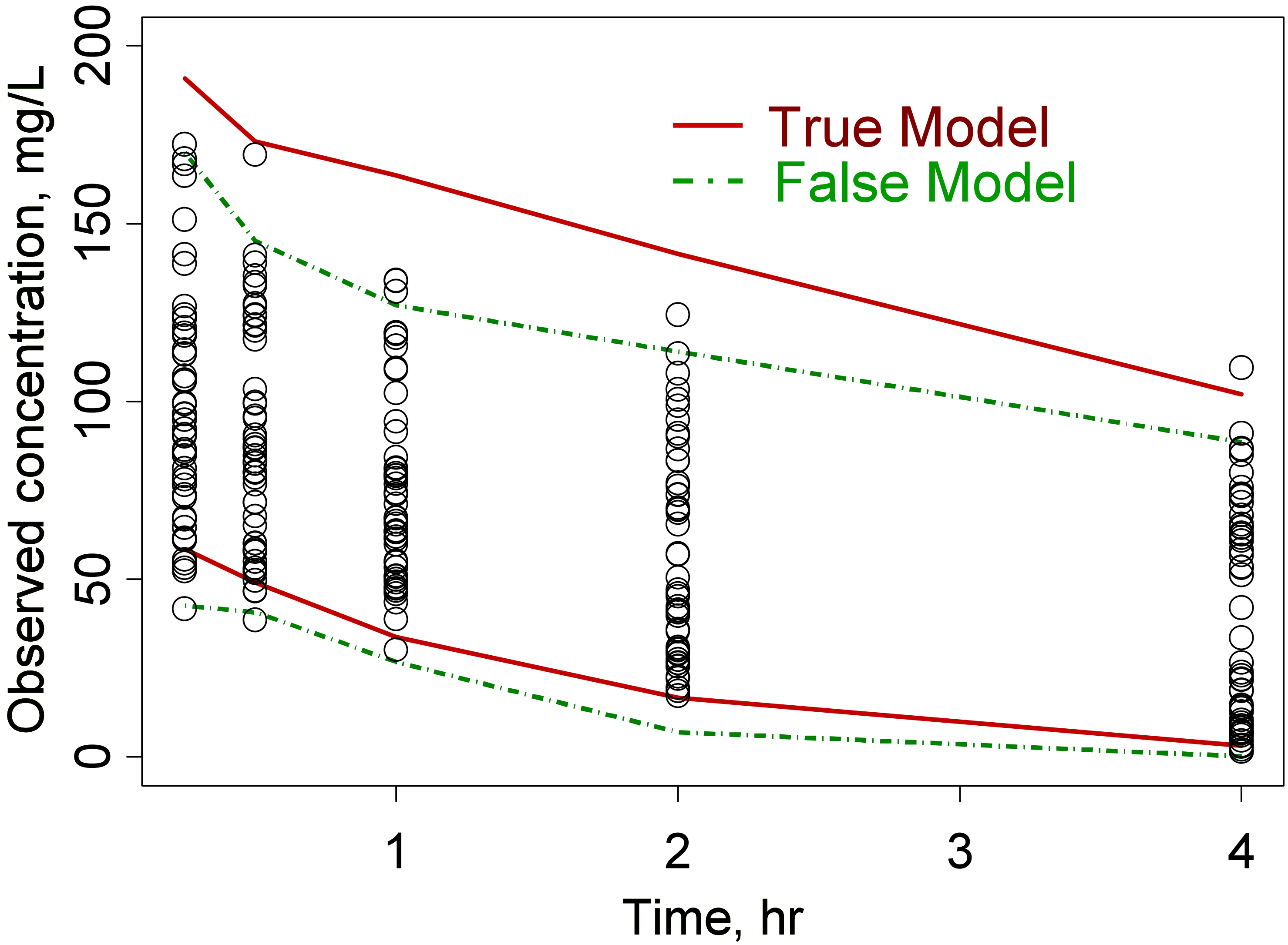
True model



? Model

False model

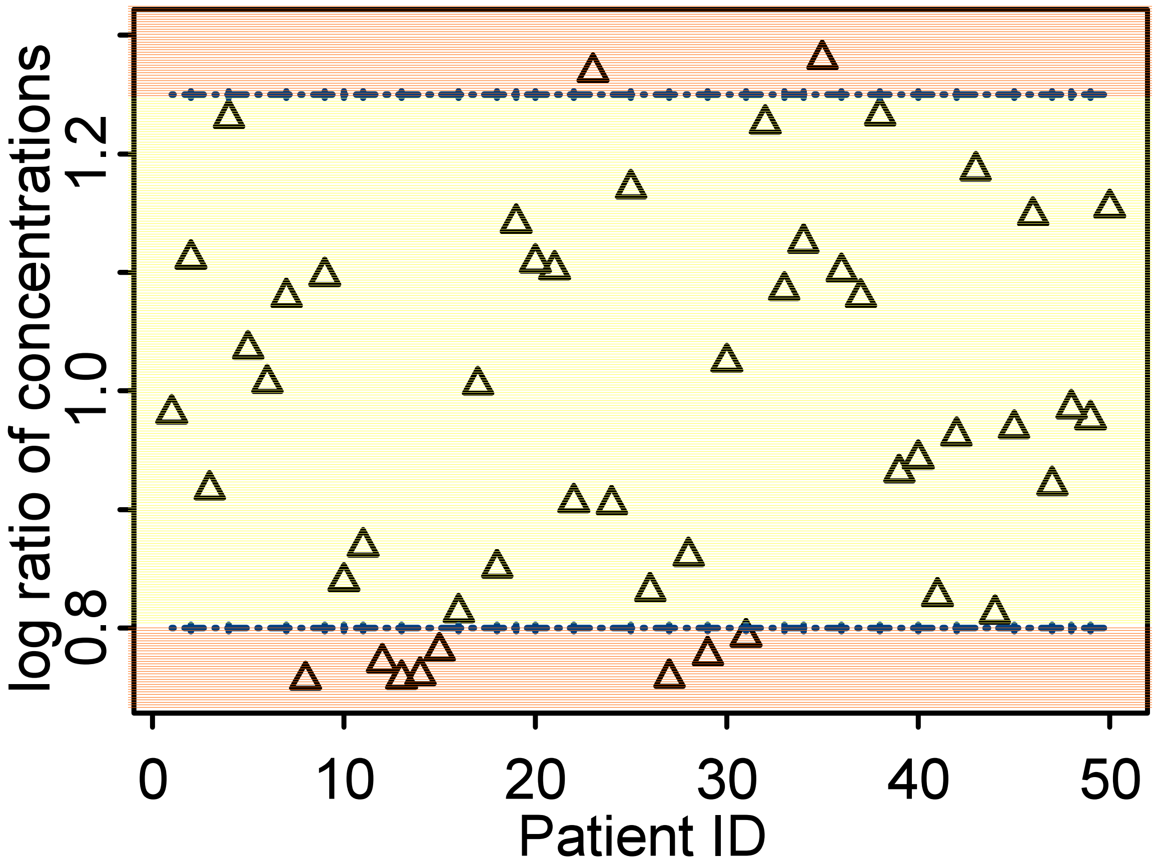
95% Prediction interval



Probability of Equivalence (p^{eqv})

- Calculation of Pp using test statistic [C(t)] would not be practically relevant.
  - e.g.: How extreme is the observed C(t) relative to replications?
- Equivalence (p^{eqv}) between observed and simulated statistic would be more relevant.

Log ratio of concentrations for p^{eqv}



Rep	ID	I()	$I()_{peqv}$
1	1	0	
1	...	...	
1	50	1	
	$\overline{I()}$=	0.8	1
2	1	0	
...	...	...	
200	50		
		0.7	0

$\overline{I()_{peqv}} = p^{eqv}$

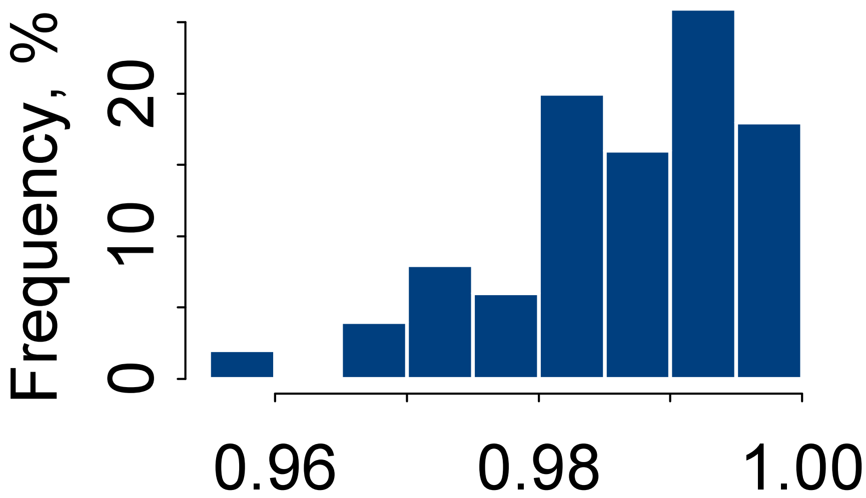
$$p^{eqv} = \frac{1}{N_{rep}} \sum_{i=1}^{N_{rep}} I\left(\frac{1}{N_{sub}} \sum_{j=1}^{N_{sub}} I\left(0.80 \le \frac{T(y_{ij}^{rep})}{T(y_j^{obs})} \le 1.25\right) > 0.75\right)$$

Probability of Equivalence (p^{eqv})

- Ideal value of p^{eqv} is 1
- Low p^{eqv} flags a poor model feature
  - acceptability is subjective, but informed
- We use two test statistics
  - C (t=0.5 hr)
  - C (t=4 hr)

Probability of Equivalence (p^{eqv})

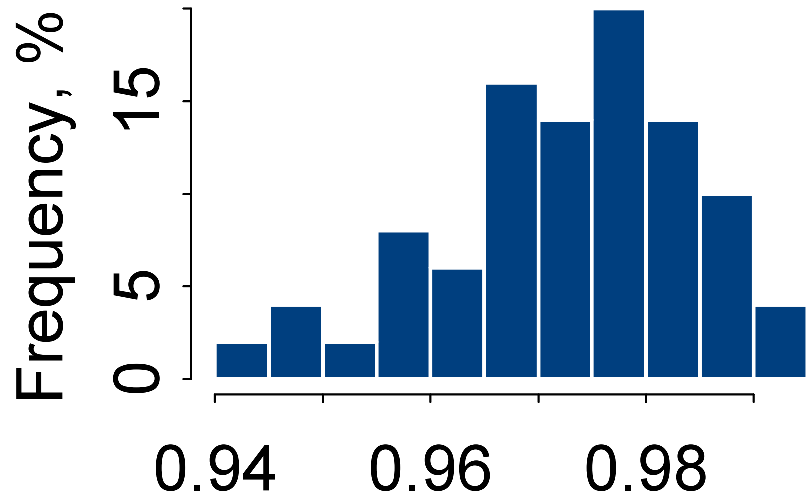
0.5 hr



p^{eqv}

? Model

True model



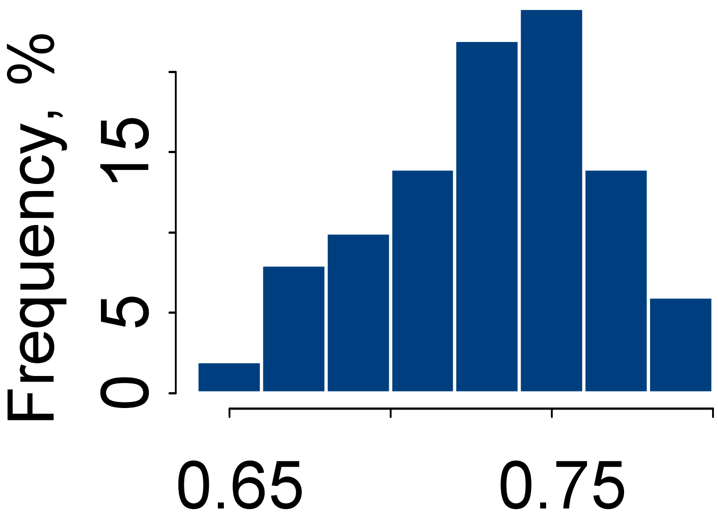
p^{eqv}

? Model

False model

Probability of Equivalence (p^{eqv})

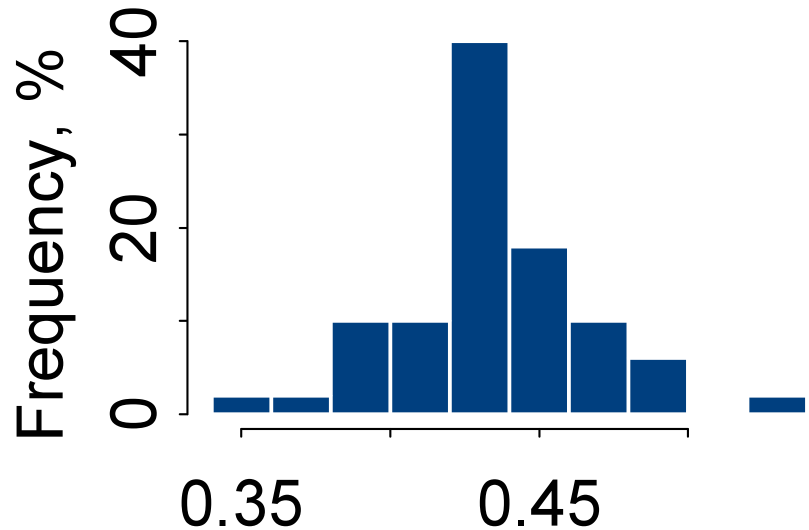
4 hr



p^{eqv}

? Model

True model



p^{eqv}

? Model

False model

Practical utility

- Simplest case was presented. Real life models are more complex.
  - **How does one decide which part of the data is important?**
  - **Selection of informative test statistic is a challenge!**

- If some how one defends the choice of domain,
  - **For the given example, we say, 20% variation (0.8-1.25) and p^{eqv} of 0.75 is acceptable.**
  - **How to set these limits?**
  - **Are we going to reject the model, if set criteria are 0.7-1.43 and most observations are JUST outside the limit?**
  - **Would the decision be different if we don't perform predictive check and rely on traditional diagnostics? Not answered!**
  - **Visual inspection would be always useful.**

Applications of PPC

- Confirming whether or not a posited model helps to describe domain specific quantity of interest
  - For example, C_{max}, C_{min}, AUC
- When a model is used for designing trials
  - Model should at least regenerate the data
- When informative priors are used (& data are spare)
  - Gelman et al
    - Reaction times in (non)schizophrenics (Bayesian)
  - Gobburu et al
    - Influence of age on renal CL in neonates (Bayesian)

Inferences

- Use of discrepancy variable cannot aid in rejecting false models.
- Use of test statistic can aid in rejecting false models.
  - Selection of informative test statistic is challenging
- 95% prediction interval cannot aid in rejecting false models.
- Equivalence based comparison of test statistic is more informative than significance based comparison.
- Further research is needed before regular use of predictive check in drug development.

References

1. Guttman I. The use of the concept of a future observation in goodness-of-fit problems. *J.R.Statist.Soc. B, 83-100, 1967.*
2. Rubin DB. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann.Statist.12,* 1151-1172*,* 1984.
3. Meng X-L. Posterior predictive p-values. *Ann. Statist.*,22, 1142-1160, 1994.
4. Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica,* 6:733-807, 1996.
5. Gobburu, Holford et al. Simplified posterior predictive check (sPPC) for model qualification. *ASCPT*, 2001.
6. Yano, Beal and Sheiner. Evaluating pk/pd models using the posterior predictive check. *J.PKPD*, 8:171-192, 2001.

Acknowledgements

- Dr. Jogarao Gobburu, OCPB/CDER/FDA
- Dr. Douglas Boudinot, VCU
- Dr. John Lawrence, Biometrics/CDER/FDA
- ECPAG organizing committee

Thank you all !!!

QUESTIONS?

Email: jadhavpr@vcu.edu