

# STRUCTURAL MODEL SELECTION

## Is the chi-square distribution appropriate for likelihood ratio testing?



Undine Falkenhagen<sup>1,2</sup>, Charlotte Kloft<sup>3</sup>, Wilhelm Huisinga<sup>2</sup>

<sup>1</sup>PharMetrX Graduate Research Training Program: Pharmacometrics & Computational Disease Modelling,

<sup>2</sup>Institut für Mathematik, Universität Potsdam, <sup>3</sup>Institut für Pharmazie, Freie Universität Berlin



### Where is the LRT with the $\chi^2$ -distribution assumption used?

- likelihood ratio test (LRT): decision between models using ratio of their likelihoods
- likelihood ratio test statistic (LRTS):

$$\Lambda(x) = 2 \log \left( \frac{\max_{\text{full model}} \ell(x|\theta)}{\max_{\text{reduced model}} \ell(x|\theta_0)} \right),$$

or equivalently the difference between their log-likelihoods

- decision based on difference in objective function value is also a likelihood ratio test:

$$\Lambda(x) = \text{OFV}_{\text{reduced model}}(x) - \text{OFV}_{\text{full model}}(x)$$

- model selection: as threshold for LRTS, the  $(1-\alpha)$ -quantile of  $\chi^2$ -distribution with  $\text{dof} = \Delta \text{number of parameters}$  is recommended (e.g. in NONMEM [1])

if  $\Lambda(x) > q_{1-\alpha}(\chi^2_p)$ , then the reduced model is rejected

- common misconception: LRTS under reduced model is asymptotically  $\chi^2$ -distributed

### Why is the $\chi^2$ -distribution used and why is this a problem?

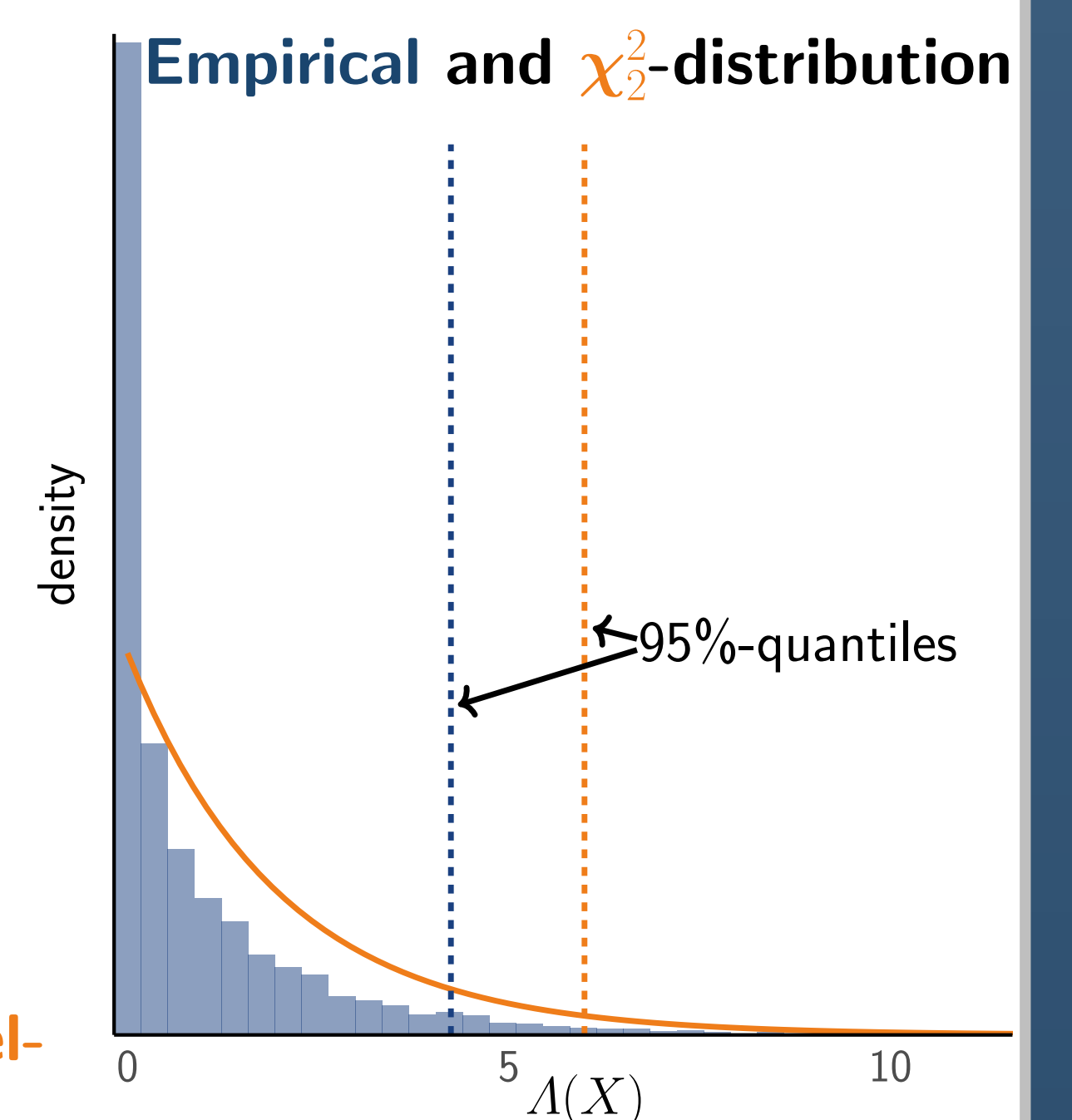
- $(1-\alpha)$ -quantile of distribution of LRTS under the reduced model as threshold guarantees asymptotic type I error  $\alpha$

**Wilks theorem [2]:** If

- models are nested
- parameters of reduced model **not on the boundary** of full parameter space
- models are **identifiable**
- likelihood functions are differentiable
- full model has  $p$  more parameters than reduced model

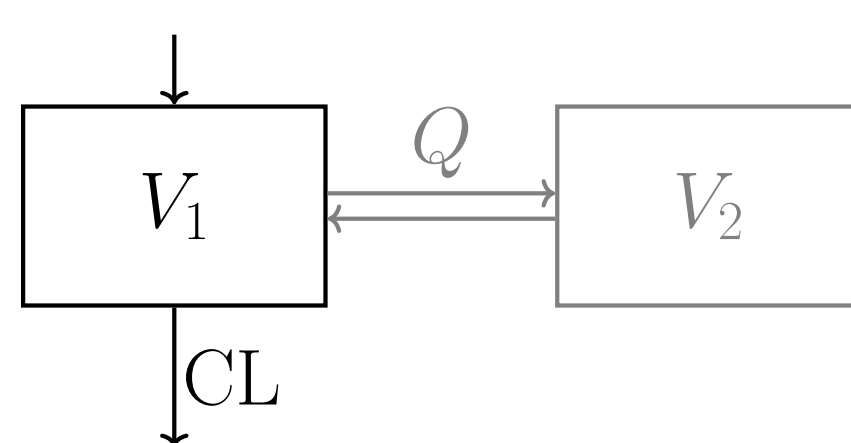
$\Rightarrow$  the LRTS  $\Lambda(x)$  under the reduced model is asymptotically  $\chi^2_p$ -distributed

**BUT: conditions not fulfilled in PK modelling** and LRTS under reduced model not  $\chi^2$



### Simulation study for decision between one- and two-compartment models

#### Cases in which a two-compartment model reduces to a one-compartment model



- $Q \rightarrow 0$ ,
- $Q \rightarrow \infty$  or
- $V_2 \rightarrow 0$

#### Workflow for parametric bootstrap

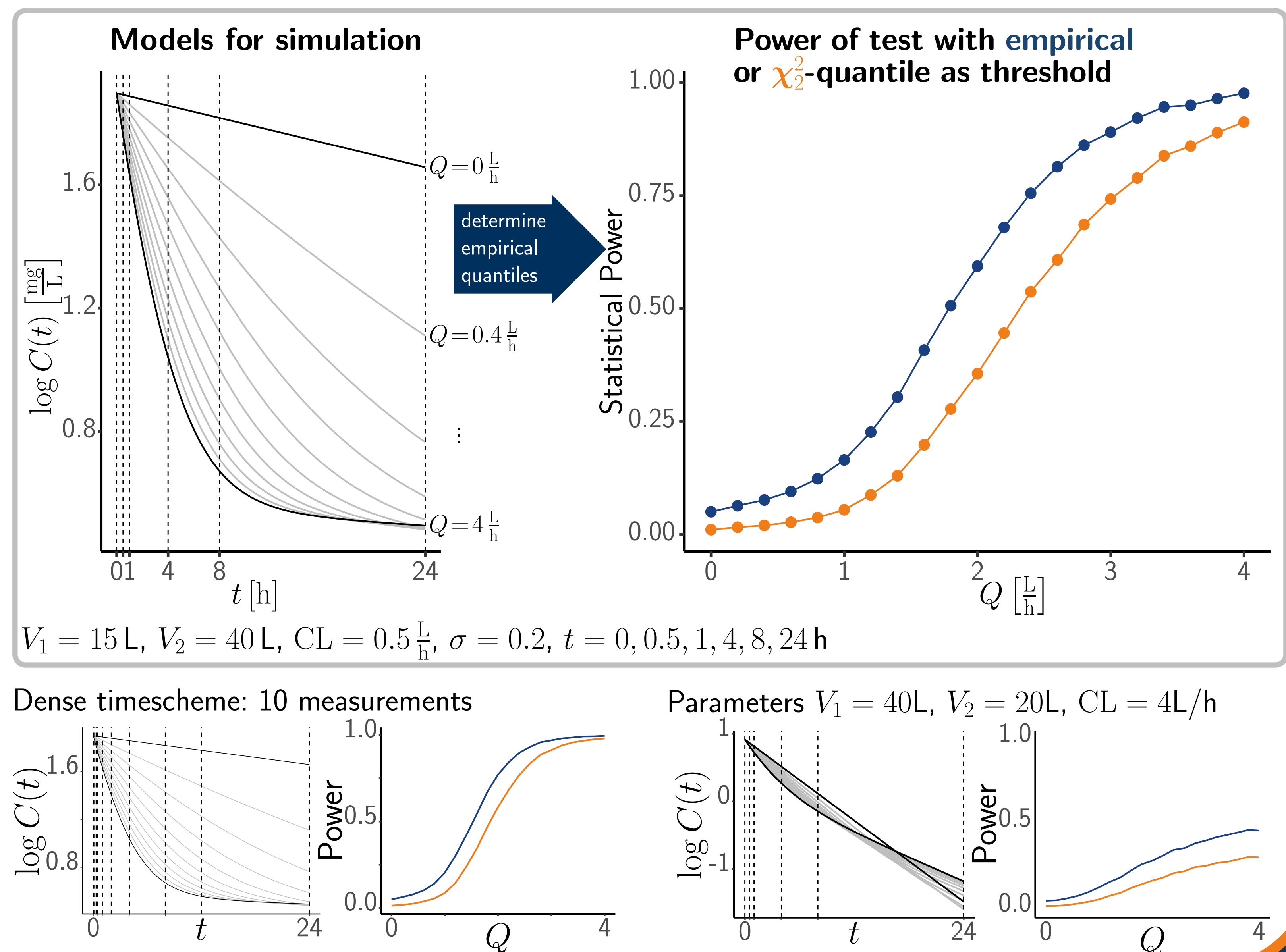
1. Fit one-compartment model to data  $x$
2. Simulate new sample data  $x_i$  from fitted model and calculate likelihood ratio test statistics  $\Lambda(x_i)$
3. Determine  $(1-\alpha)$ -quantile  $q_{1-\alpha}(\Lambda(x_i))$  of test statistics
4. Calculate test statistic  $\Lambda(x)$  for data  $x$
5. Classify data as two-compartment system if  $\Lambda(x) > q_{1-\alpha}(\Lambda(x_i))$

#### How to estimate two-compartment models with barely identifiable parameters to prevent numerical problems due to unidentifiability:

1.  $V_2$  fixed on a discrete grid (uniform on log-scale) and  $V_1$ ,  $Q$ ,  $CL$  and residual error  $\sigma$  estimated
2. all parameters continuously estimated in neighbourhood of best model estimates from previous step

#### Power comparison dependent on threshold

- classical one- and two-compartment models with multiplicative log-normal residual error
- power increases with more distinct deviation from one-compartment model (different  $Q$  values)



### What has been done previously?

Work on asymptotics for likelihood ratio tests under non-identifiability:

- LRTS under the reduced model is asymptotically the supremum of a squared Gaussian process (quantiles also typically determined by simulation) [3]
- for application on one- and two-compartment systems assumptions like uniform distribution of sampling time needed (not realistic) [4]

Work on parametric bootstrap in pharmacometric modelling:

- in covariate modelling,  $\chi^2$  seems to be anti-conservative, i.e. when using  $\chi^2$ -quantiles, the nominal  $\alpha$ -level is not attained [5]

### How does the $\chi^2$ -quantile as threshold influence model selection?

In our simulations:

- 95%-quantile of  $\chi^2$  up to twice as large as actual 95%-quantiles
- aiming for  $\alpha$ -level of 5% results in actual  $\alpha$ -level of  $\sim 1$ -2%
- $\chi^2$ -quantiles are more restrictive than intended  $\rightarrow$  reduced power of the test

**When using  $\chi^2$ -quantiles, contrary to what is intended, no statements can be made about the level of the test and it can be far from the targeted value.**

### References

- [1] Lewis B. Sheiner. Analysis of pharmacokinetic data using parametric models. III. Hypothesis tests and confidence intervals. *J Pharmacokinet Biopharm*, 14(5):539–555, oct 1986.
- [2] Samuel S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1):60–62, mar 1938.
- [3] Yongzhao Shao and Xin Liu. Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.*, 31(3):807–832, jun 2003.
- [4] Tobias Machewitz. Likelihood-Ratio-Test bei nicht identifizierbaren Parametern. Master's thesis, Universität Potsdam, 2016.
- [5] Hoai-Thu Thai, France Mentré, Nicholas H. G. Holford, Christine Veyrat-Follet, and Emmanuelle Comets. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. *J Pharmacokinet Pharmacodyn*, 41(1):15–33, dec 2013.