Boehringer Ingelheim

Workshop of the Bayes WG / IBS-DR
Mainz, 2006-12-01

G. Nehmiz
M. Könen-Bergmann

Model validation through "Posterior predictive checking" and "Leave-one-out"

Overview

The posterior predictive distribution from a fitted model

Check of fit between model and data

The “Leave-one-out” method for the 1-way ANOVA model

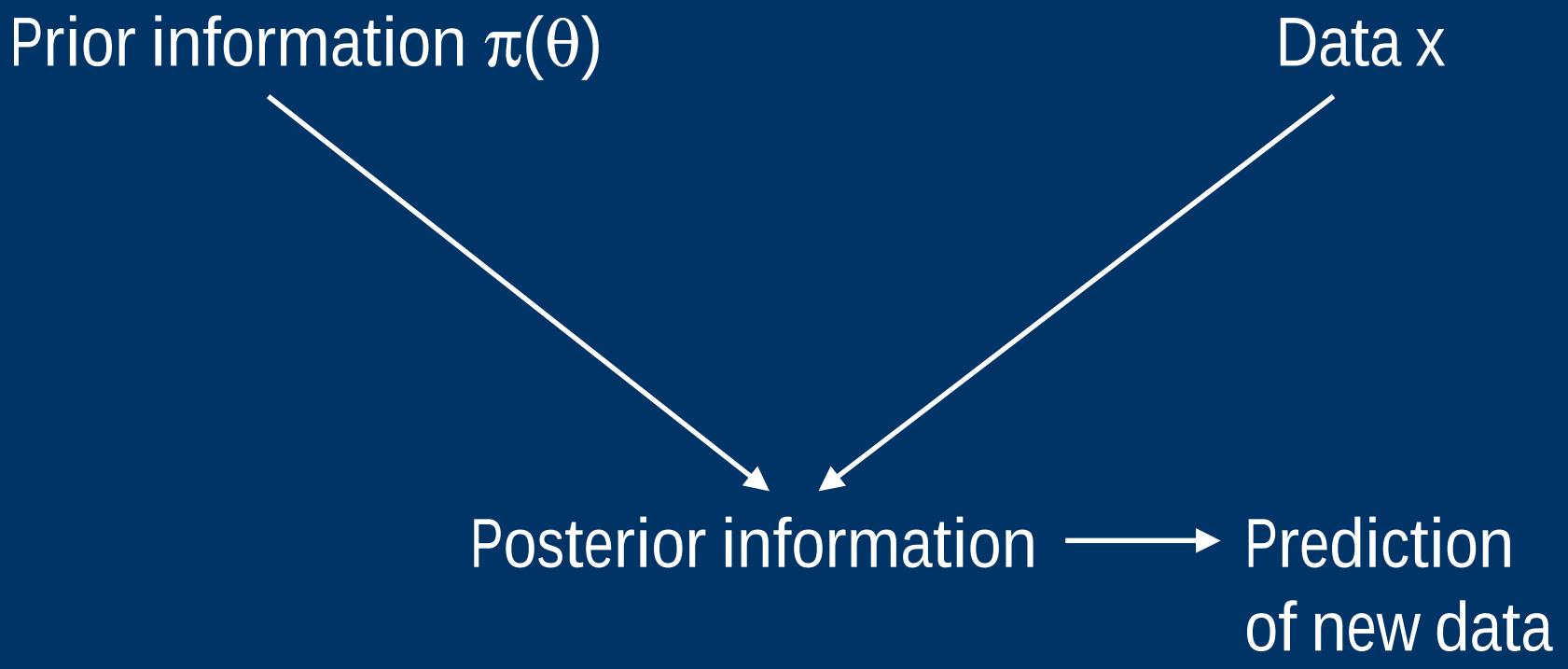
Example: ECG data

Summary

References

The posterior predictive distribution from a fitted model

Boehringer Ingelheim



Information is represented by probability distributions on the parameter space Θ

The posterior predictive distribution from a fitted model

Boehringer Ingelheim

Probability model: p(x|θ)

$$p(J \mid x) = \frac{p(J) \cdot l(J \mid x)}{\int_{\Theta} p(J) \cdot l(J \mid x) \, dJ}$$

posterior distribution

(norm. factor)

$$p_p(\tilde{x} \mid x) = \int_{\Theta} p(\tilde{x} \mid J) \cdot p(J \mid x) \, dJ$$

Predictive distribution for new data

Check of fit between model and data

Boehringer Ingelheim

Model selection – comparison of = 2 models with each other

Model validation – consideration of 1 model and of its fit to the data, without reference to (an) alternative model(s)

We are now concerned with model validation only.

Check of fit between model and data

Boehringer Ingelheim

Data prediction as a means of model validation: Subdivide data into learning sample and validation sample, and compare the data of the validation sample with the values predicted from the learning sample (better: predicted from the posterior distribution derived from the learning sample).

Check of fit between model and data

Boehringer Ingelheim

(a) Learning sample and validation sample both of considerable size – difficult to investigate

(b) Learning sample empty – predict all data from the prior distribution (“prior predictive check”)

Check of fit between model and data

Boehringer Ingelheim

(c) Validation sample empty – fit model to all data and re-check ("posterior predictive check"). Values predicted from $\pi(\theta|x)$ will the formally not be "new" data but only replicates of the observed data (all covariate values remain the same).
See Gelman/Carlin/Stern/Rubin 2004, O'Hagan 2003.

(d) Leave-one-out method – predict each data point x_i from the posterior distribution derived from all others, $\pi(\theta|x_{-i})$.

Check of fit between model and data

Boehringer Ingelheim

If $\pi(\theta|x)$ or $\pi(\theta|x_{-i})$ is determined by MCMC simulation, predicted values can be generated at each iteration and the distribution of these predicted values can be compared with the data point x_i itself

The aberrant position of x_i relative to the distribution of the predicted values is described by the "predictive p-value" $P(x_i\tilde{} = x_i \mid x)$ or $P(x_i\tilde{} = x_i \mid x_{-i})$

Check of fit between model and data

Boehringer Ingelheim

Predictive p-values close to 0.5 show that the fit for that data point is good

Calibration is a difficult problem: Which deviation from 0.5 should be considered as a relevant lack of fit? Predictive p-values are not U[0,1] distributed, see e.g. Hjort/Dahl/Steinbakk (2006). Remains open for artificial data (O'Hagan 2003, Sharples 1990). Therefore we turn to measured data (ECG data) where an external (medical) relevance assessment exists

We investigate now methods (c) and (d). Method (c), based on $\pi(\theta|x)$, is simpler – but is it adequate?

The “Leave-one-out” method for the 1-way ANOVA model

Boehringer Ingelheim

$$x_{ij} = m_i + s \cdot e_{ij}$$ with ε_{ij} i.i.d. N(0,1) and common σ

$$m_i = m + t \cdot e_i$$ with ε_i i.i.d. N(0,1) and independent of the ε_{ij}

The “Leave-one-out” method for the 1-way ANOVA model

Boehringer Ingelheim

- Marshall/Spiegelhalter (2003) investigate analytically the balanced case with known σ and τ
- The degree of overoptimism of the posterior predictive check depends from I, the number of groups, and decreases with increasing I
- Also, they propose a “Leave-one-group-out” method

The "Leave-one-out" method for the 1-way ANOVA model

Boehringer Ingelheim

$$x_{ij} = m_i + s \cdot e_{ij}$$ with ε_{ij} i.i.d. N(0,1) and common σ

$$m_i = m + t \cdot e_i$$ with ε_i i.i.d. N(0,1) and independent of the ε_{ij}

Prior distributions:

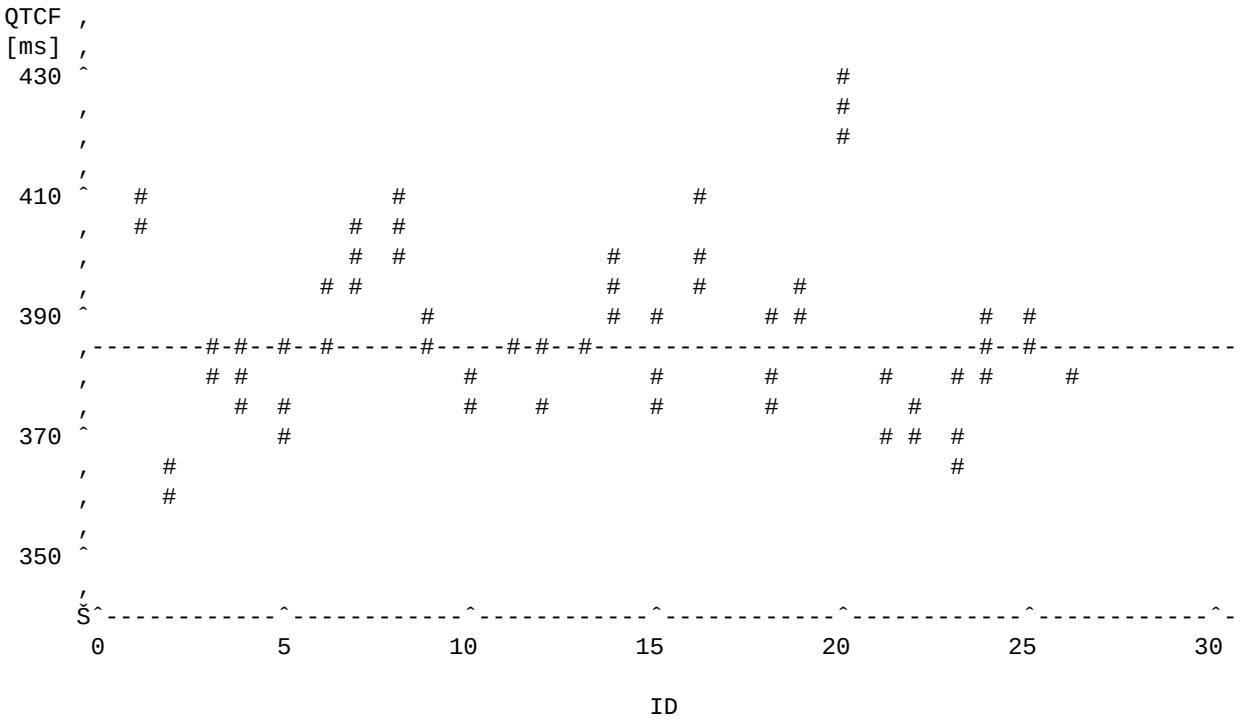
$\sigma \sim U(0,S)$ with S large

$\tau \sim U(0,T)$ with T large

$\mu \sim N(0,U)$ with U large

Example: ECG data

25 subjects, 3 repetitions



Example: ECG data

Boehringer Ingelheim

Potentially critical subjects for QTcF are those with a span of at least 12 ms between repetitions, and subject 20

We investigate now subject 16 (repetition 1, value 409 ms) and 20 (all 3 repetitions)

See Camm (2006) for explanation of ECG intervals and correction methods

Example: ECG data

Hierarchical random-effects model is fitted through MCMC (see Gilks/Richardson/Spiegelhalter 1996) and formulated in WinBUGS

```
Model
{
for (l in 1:L) {
  y[l]~dnorm(mi[subj[l]],sigi);
       }
for (k in 1:K) {
  mi[k]~dnorm(mu,taui);
       }
#
y161~dnorm(mi[16],sigi);
check161 <- step(409-y161);
...
mu~dnorm(390,1.0E-4);
sigma~dunif(0,1000);
sigi <- 1/(sigma*sigma);
tau~dunif(0,1000);
taui <- 1/(tau*tau);
}
```

Example: ECG data

Boehringer Ingelheim

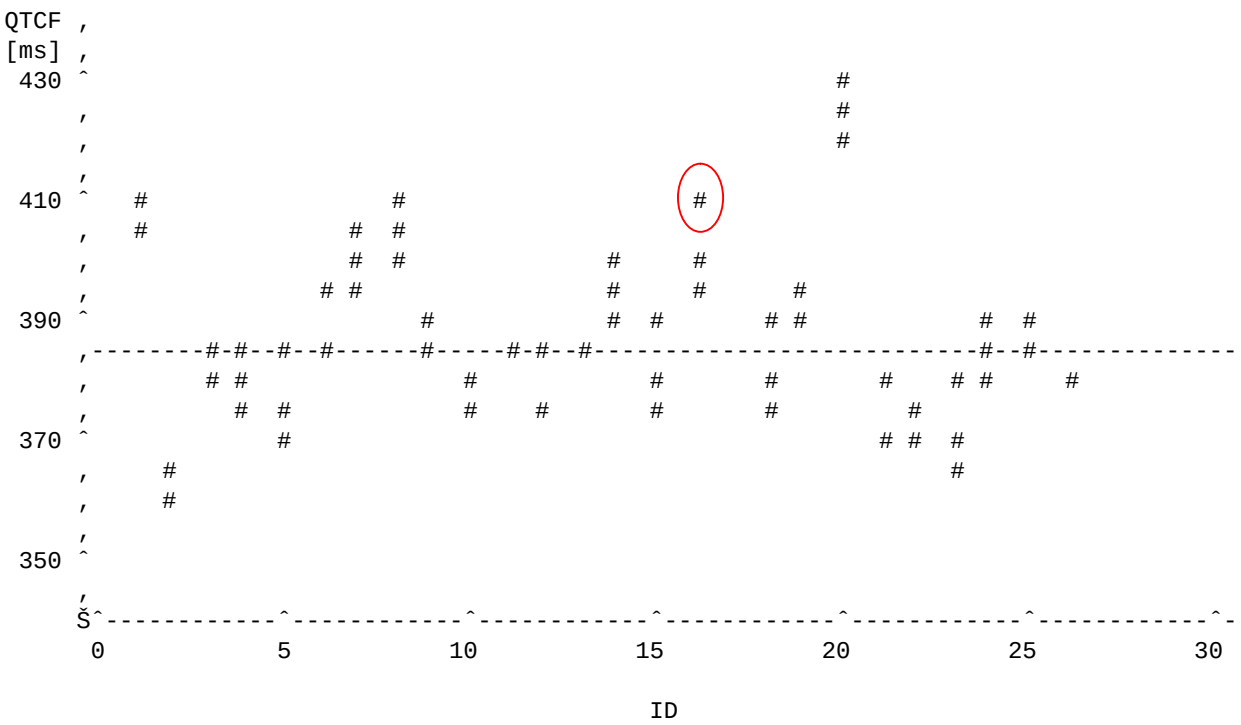
Hierarchical random-effects model is fitted through MCMC (see Gilks/Richardson/Spiegelhalter 1996) and formulated in WinBUGS

```
# .../ecg161d.txt
# Value of subject 16,
# repetition 1 (409) left out
#
list(
y=c(                                    subj=c(
403,410,408,                            1,1,1,
...                                     ...
393,400,                                16,16,
...                                     ...
382,381,381),                           25,25,25),
                                        K=25,L=74)
```

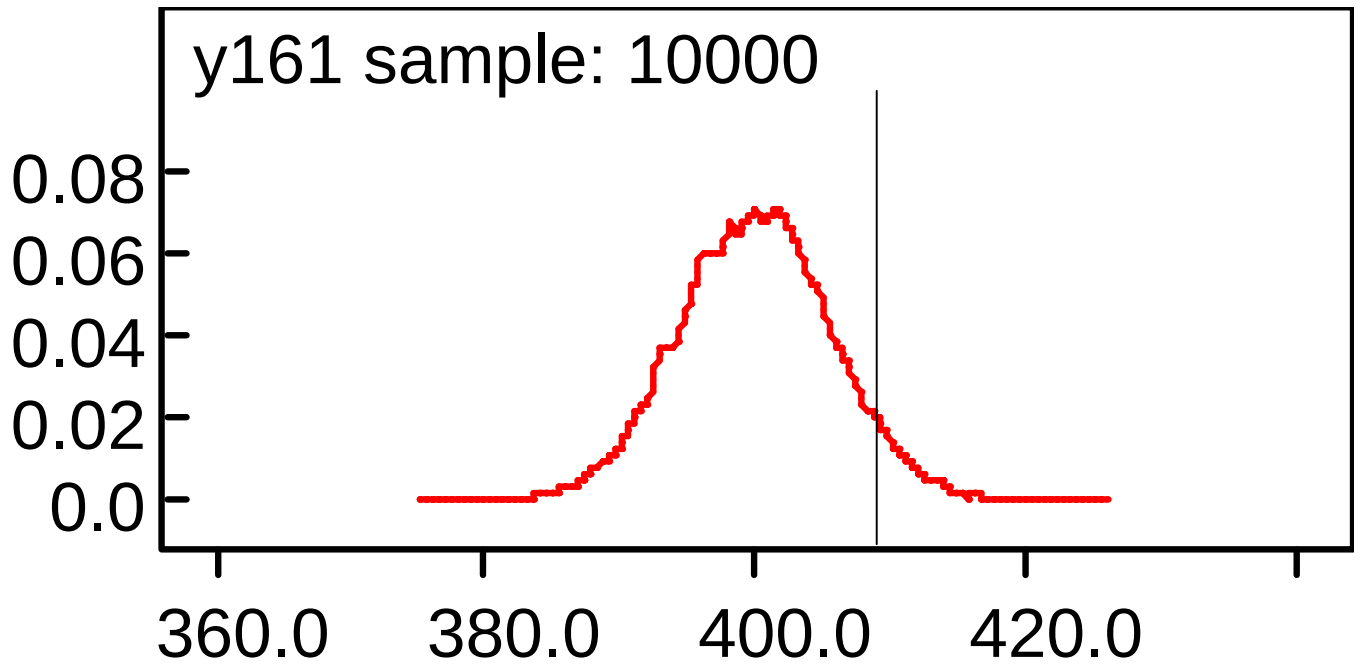
Example: ECG data with model fitted to all data

Boehringer Ingelheim

25 subjects, 3 repetitions



Example: ECG data with model fitted to all data



Example: ECG data with model fitted to all data

Boehringer Ingelheim

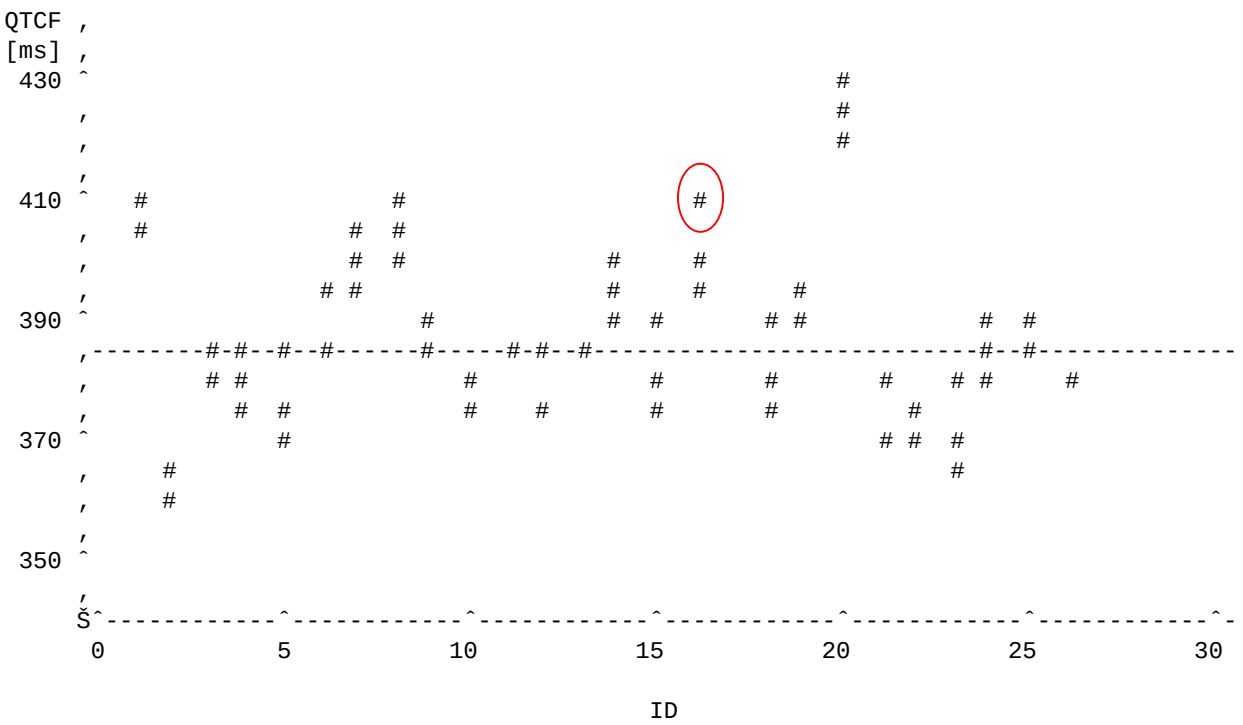
Repetition 1 of subject 16 (measured: 409 ms) is predicted as 400 +/- 5.7 ms

The predictive p-value for $x_{16,1}$ is 0.9427

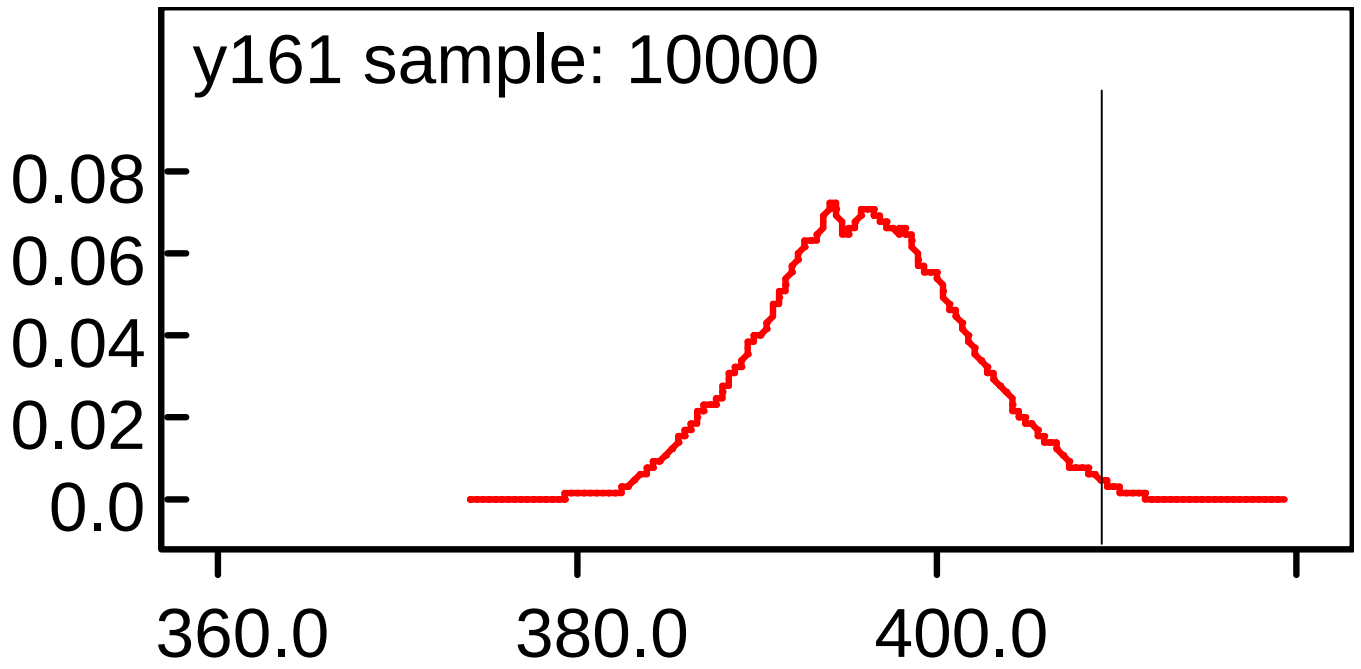
Example: ECG data without data point in question

Boehringer Ingelheim

25 subjects, 3 repetitions



Example: ECG data without data point in question



Example: ECG data without data point in question

Boehringer Ingelheim

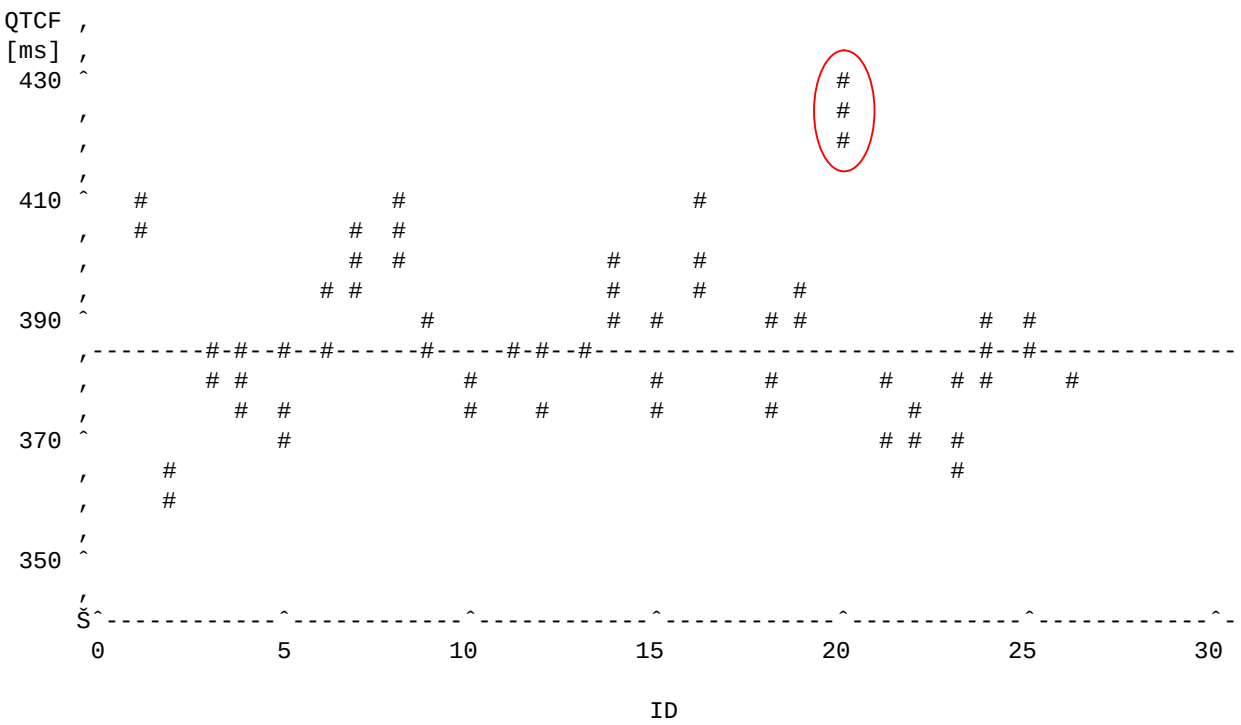
Repetition 1 of subject 16 (measured: 409 ms) is predicted as 396 +/- 5.7 ms

The predictive p-value for $x_{16,1}$ is 0.9896

Example: ECG data without data point in question

Boehringer Ingelheim

25 subjects, 3 repetitions



Example: ECG data without data point in question

Omission of any single data point of subject 20 does not change the situation as the 3 points “mask” each other

Example: ECG data without data point in question

Boehringer Ingelheim

Medically, also the data point with a “leave-one-out” predictive p-value of 0.9896 is explainable by biological variation in healthy volunteers – short peak of HR without adaptation of QT (“QT hysteresis”) is a plausible explanation

Think of an extended model with individual σ_is

Summary

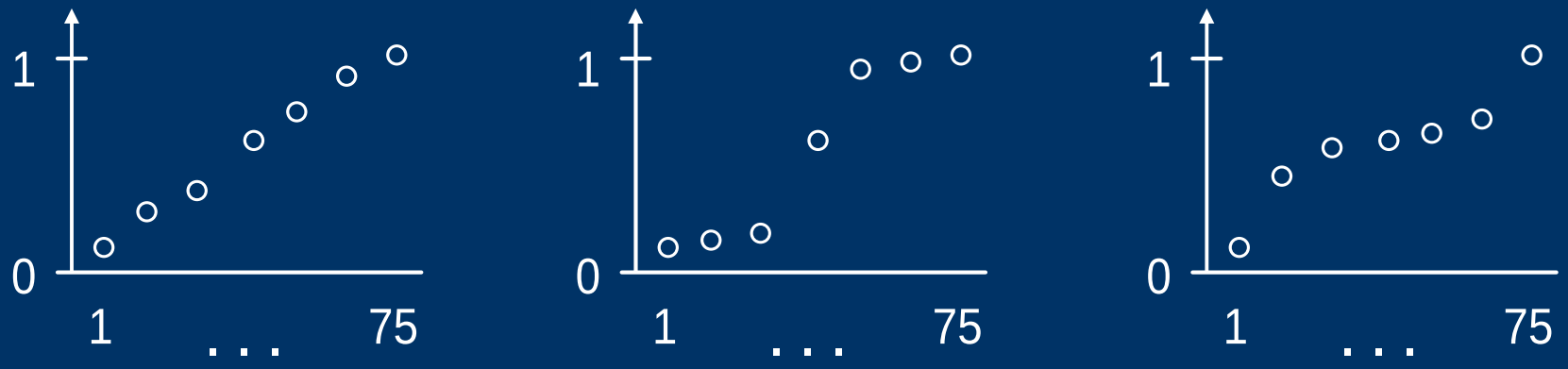
Boehringer Ingelheim

"Leave-one-out" is sensitive to masking (not new)

Comparing the data with the values predicted from the model fitted to all data (including the data point in question) is overoptimistic, predictive p-values are pulled towards 0.5

Summary

Calibration remains a problem for each model validation procedure



References

Gelman A, Carlin JB, Stern HS, Rubin DB:
Bayesian Data Analysis (2. ed.).
Boca Raton / London / New York / Washington/DC: Chapman & Hall / CRC 2004.

O'Hagan A:
HSSS model criticism. In:
Green PJ, Hjort NL, Richardson S (eds.):
Highly Structured Stochastic Systems.
Oxford: Oxford University Press 2003, 423-453.

Hjort NL, Dahl FA, Steinbakk GH:
Post-Processing Posterior Predictive p Values.
Journal of the American Statistical Association 2006; 101 (475): 1157-1174.

Sharples LD:
Identification and accommodation of outliers in general hierarchical models.
Biometrika 1990; 77 (3): 445-453.

References

Marshall EC, Spiegelhalter DJ:
Approximate cross-validatory predictive checks in disease mapping models.
Statistics in Medicine 2003; 22: 1649-1660.

Camm AJ:
How does pure heart rate lowering impact on cardiac tolerability?
European Heart Journal Supplements 2006; 8 (D): D9-D15.

Gilks WR, Richardson S, Spiegelhalter DJ (Hg.):
Markov Chain Monte Carlo in Practice.
London / Weinheim / New York / Tokyo / Melbourne / Madras: Chapman & Hall 1996.