Bayesian Coefficients of Variation in Linear Mixed Models,

Random Effects and Precision in Assay Qualification

Bernard G Francq
Dan Lin
Mikaël Le Bouter
Mélissa Scozzari
Walter Hoyer

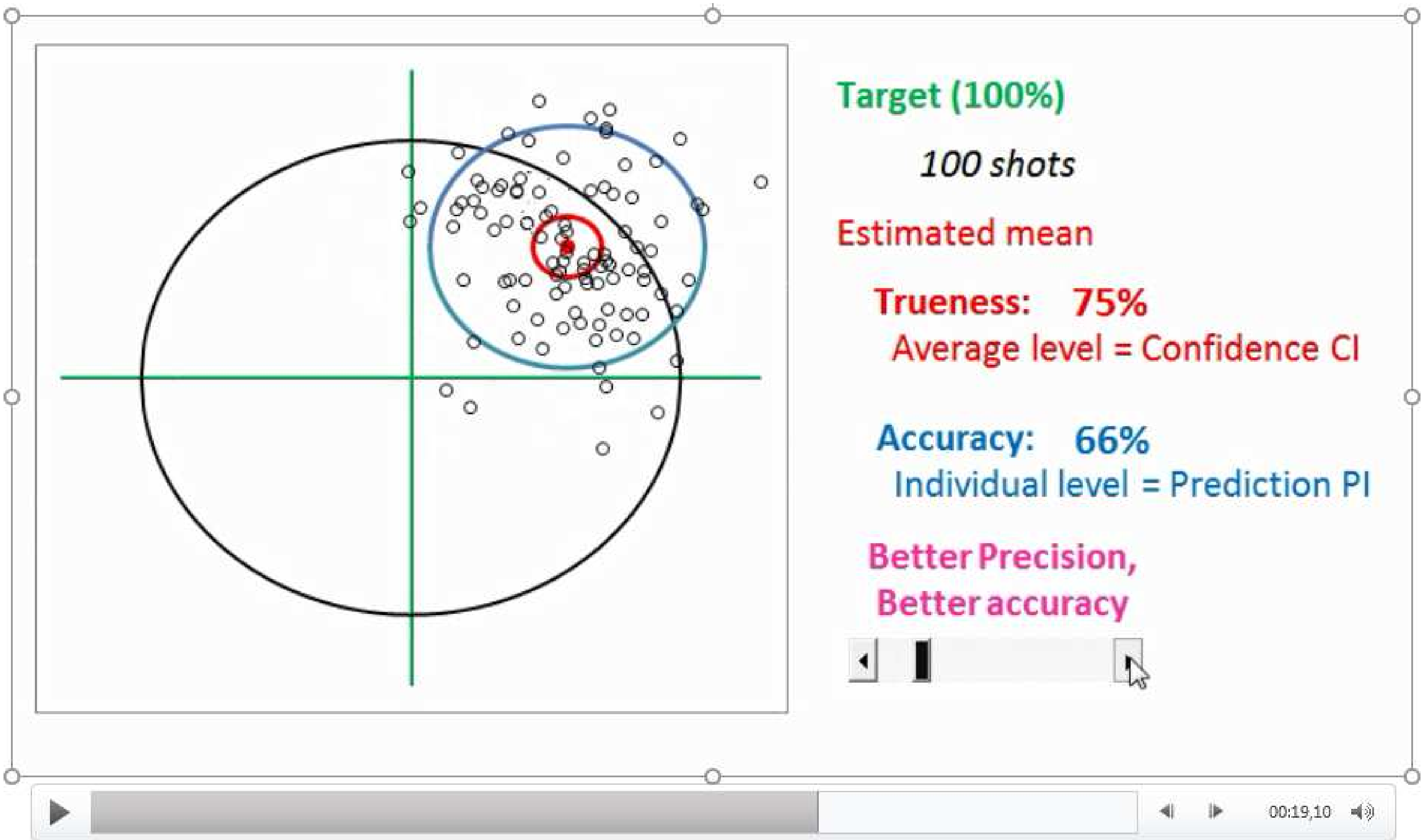
BAYES BAYESIAN BIOSTATISTICS LYON 2019 21-24 MAY

TUESDAY 21ST FRIDAY 24TH MAY 2019
MARRIOTT HOTEL CITÉ INTERNATIONALE, LYON, FRANCE

Measurement Science

What we expect from a measure:

Precise + True = Accurate



When I play drums *precisely* and *accurately*…

Introduction to Method Qualification (and Validation)

Aim of qualification

- the analytical method is suitable for its intended use
- and consequently to prove the reliability of the results obtained

Qualification statistics considered

- **Precision**
- Trueness
- Accuracy

Experimental design

- Multiple replicates per sample
- Multiple days/operators/sessions
- Series dilutions of a spiked-in sample or known concentrations

Precision, Trueness and Accuracy

	Precision +	Trueness =	Accuracy
Meaning	Random error	Systematic Error	Total error
Related to	Method variability	Method bias	Total deviation from nominal value
Quantified by	**CV or STD**	**CI** Confidence Interval	**PI or TI** Prediction or Tolerance Interval

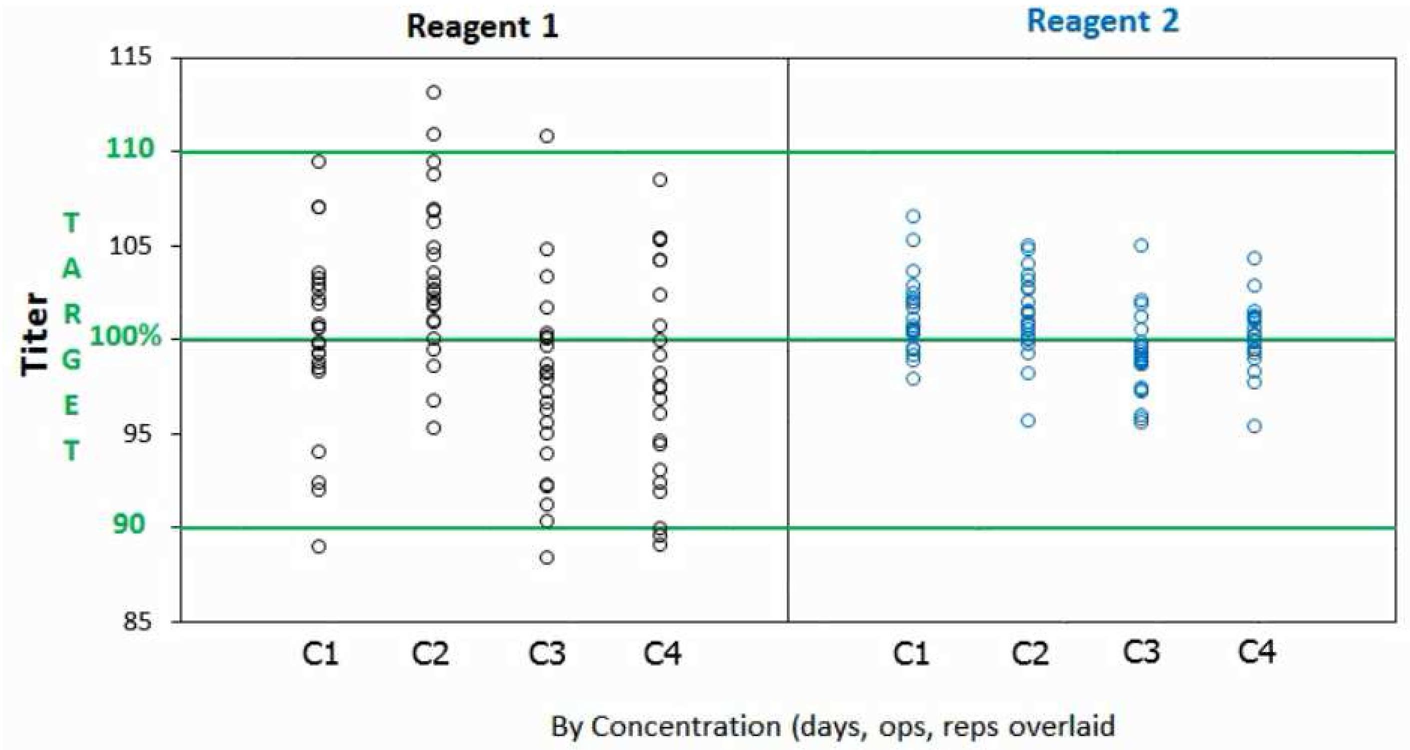
of difference to nominal value

Data Set

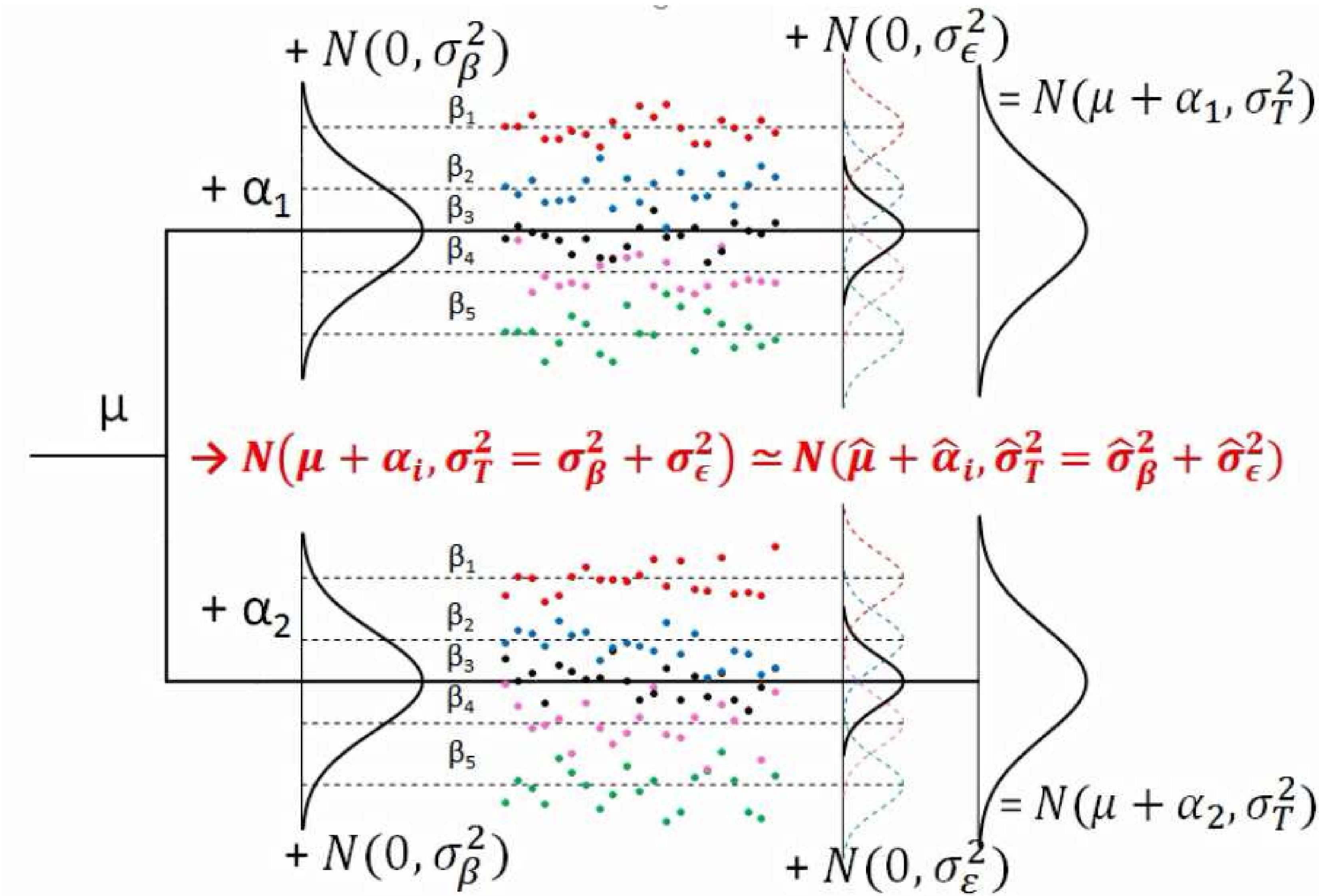
The study design for the assay validation is composed of:

- 2 different reagents(R1, R2): fixed variable
- 4 operators (B, D, S, W): random variable
- 3 days (D1, D2, D3): random variable
- 2 replicates
- 4 nominal concentrations (25, 50, 75, 100) µl: *fixed variable*

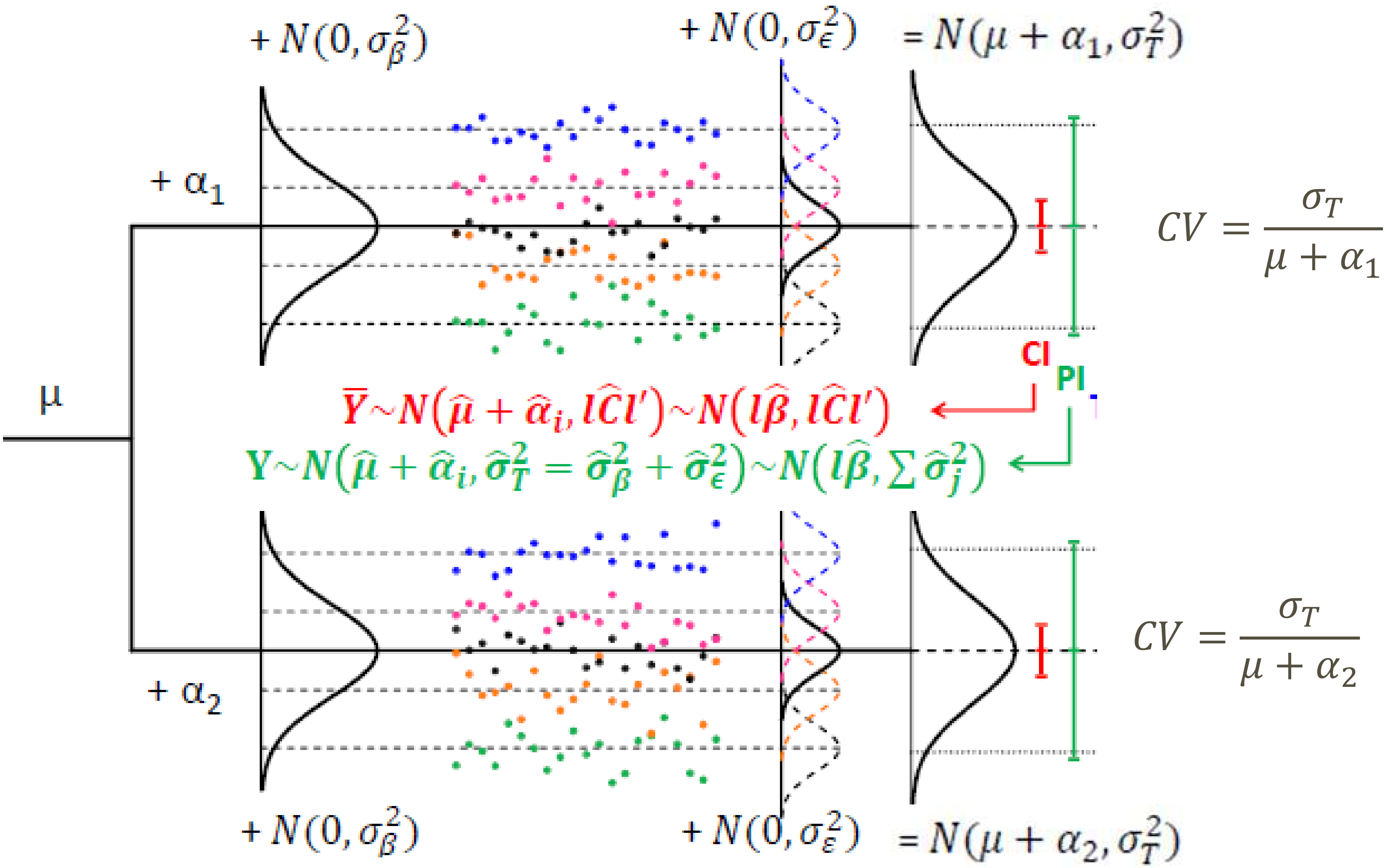
Crossed random effect (operators and days)



Mixed Model Formulation



Confidence, Prediction, CV in Mixed Model



Trueness (CI), Accuracy (PI), Precision (CV)

Confidence intervals are used to assess the trueness

- The degrees of freedom are typically calculated by Kenward-Roger (KR) method
- A plot can be displayed with the CIs calculated at the different level of concentrations

Prediction intervals are used to assess the accuracy

- An accuracy profile can be displayed with the PIs calculated at the different level of concentrations
- The uncertainty of the prediction is then the sum of the systematic error (Trueness) + random error (Precision)

Coefficients of Variations (CVs) are used to assess the precision

- The degrees of freedom are calculated by the Generalized Satterthwaite method
- **Frequentist** 95% CI are calculated from an adaptation of the modified McKay formula (for univariate distribution)
- **Bayesian** statistics is a straightforward approach to obtain posterior distribution and 95% credible or HPD intervals

Trueness and Accuracy can be expressed in percentage, as well as CV

CV and its 95% CI in univariate distribution

Under the normality assumption,

$$CV = \frac{\sigma}{\bar{\bar{X}}}$$

Its frequentist 95% CI is given by the *modified McKay formula:*

$$\frac{CV}{\sqrt{\left(\frac{\chi^2_{\kappa,r}}{r+1} - 1\right) CV^2 + \frac{\chi^2_{\kappa,r}}{r}}}$$

Where $r = n - 1$ and $\kappa = 0.025$ (or $\kappa = 0.975$) for the lower (upper) bound.

Do you know another formula to calculate the CV?

CV and its 95% CI in univariate distribution

Under log-normal data, the CV is only related to the variance (on the log scale):

$$CV = \sqrt{e^{\sigma^2} - 1}$$

Its frequentist 95% CI is given by the classical 95% CI for σ^2

(Not shown in this presentation)

CV and its 95% CI in mixed models

In mixed models,

- ✓ The CV is calculated per variance components
- ✓ Total variance = Intermediate Precision
- ✓ The mean is replaced by the fixed effects estimate (i.e. intercept)
- ✓ Under normality assumption:

$$CV_T = \frac{\sigma_T}{l\beta} \text{ estimated by } \frac{\hat{\sigma}_T}{l\hat{\beta}}$$

- ✓ Under log-normal data:

  the CV are directly related to the variance components, and the 95% CI for CV is related to the classical 95% CI for the variance components.

  (Not shown in this presentation)

CV and its 95% CI in mixed models - Frequentist

In mixed models,

$$CV_T = \frac{\sigma_T}{l\beta} \text{ and CI for CV} = \frac{CV}{\left(\frac{\chi^2_{\kappa,r}}{r+1} - 1\right) CV^2 + \frac{\chi^2_{\kappa,r}}{r}}$$

Improvement

We need a generalized formula
for a wide variety of designs in mixed models
(one random factor, nested and crossed designs
for multiple random factors, balanced or unbalanced designs)

The 95% CI for (the total) CV is calculated from an adaptation of the McKay formula with degrees of freedom by the Generalized Satterthwaite formula

CV and its 95% CI in mixed models - Bayesian

In Bayesian mixed models, the CV can be obtained from MCMC simulations, with its 95% credible or HPD intervals

PROC MCMC in SAS

1-way random (operator) model

```
PROC MCMC DATA = Set3 NBI = 10000 NMC = 10000 STATISTICS = Intervals;
PARMS B0 S2;
PARMS S2op 1;
PRIOR B0  ~ normal(0, var=1e6);
PRIOR S2 ~ igamma(0.01, scale = 0.01); or half-Cauchy distribution
prior S2op ~ igamma(0.01, scale = 0.01); or half-Cauchy distribution
random Gamma ~ normal(0, var = S2op) subject = op;
Mu = B0 + Gamma;
S2tot = S2op + S2;
cvtot = sqrt (S2 + S2op) / B0;
MODEL resp ~ normal(Mu, var = S2);
RUN;
```

Parametrization in mixed model

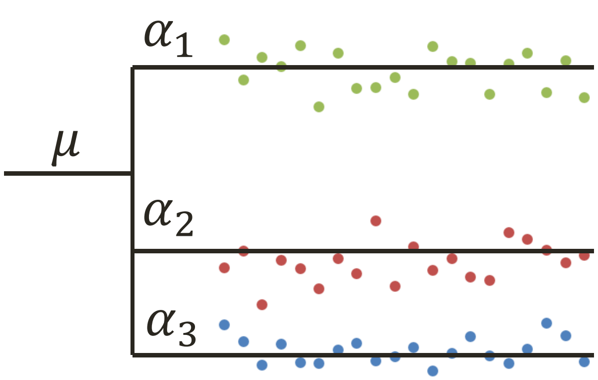
Our guidelines

Fixed effects

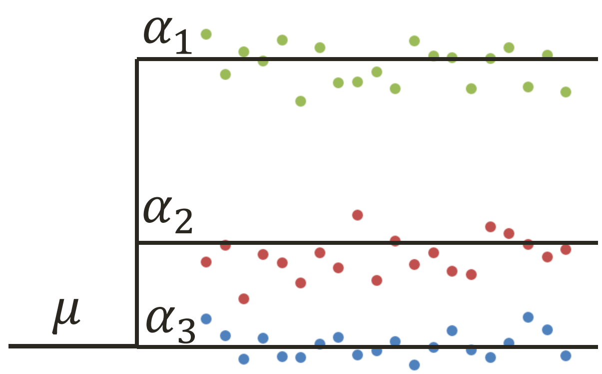
- Cell means model (no intercept)
- Combine all fixed effects into 1 variable

Random effects

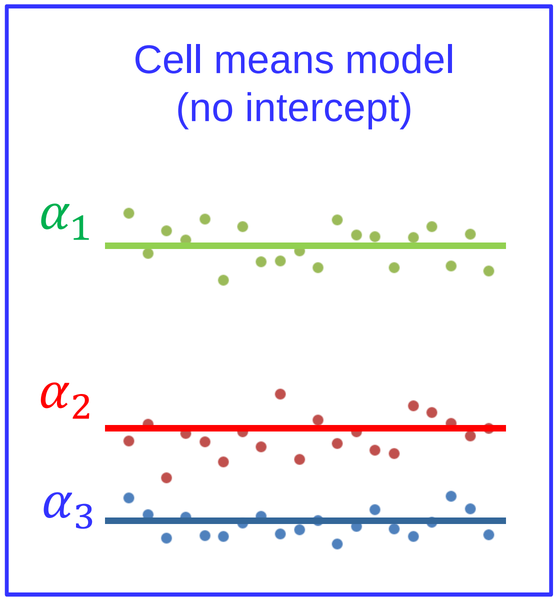
- Reflect the actual design of experiments (no simplification)
- Omitting or combining random effects can underestimate the total variance



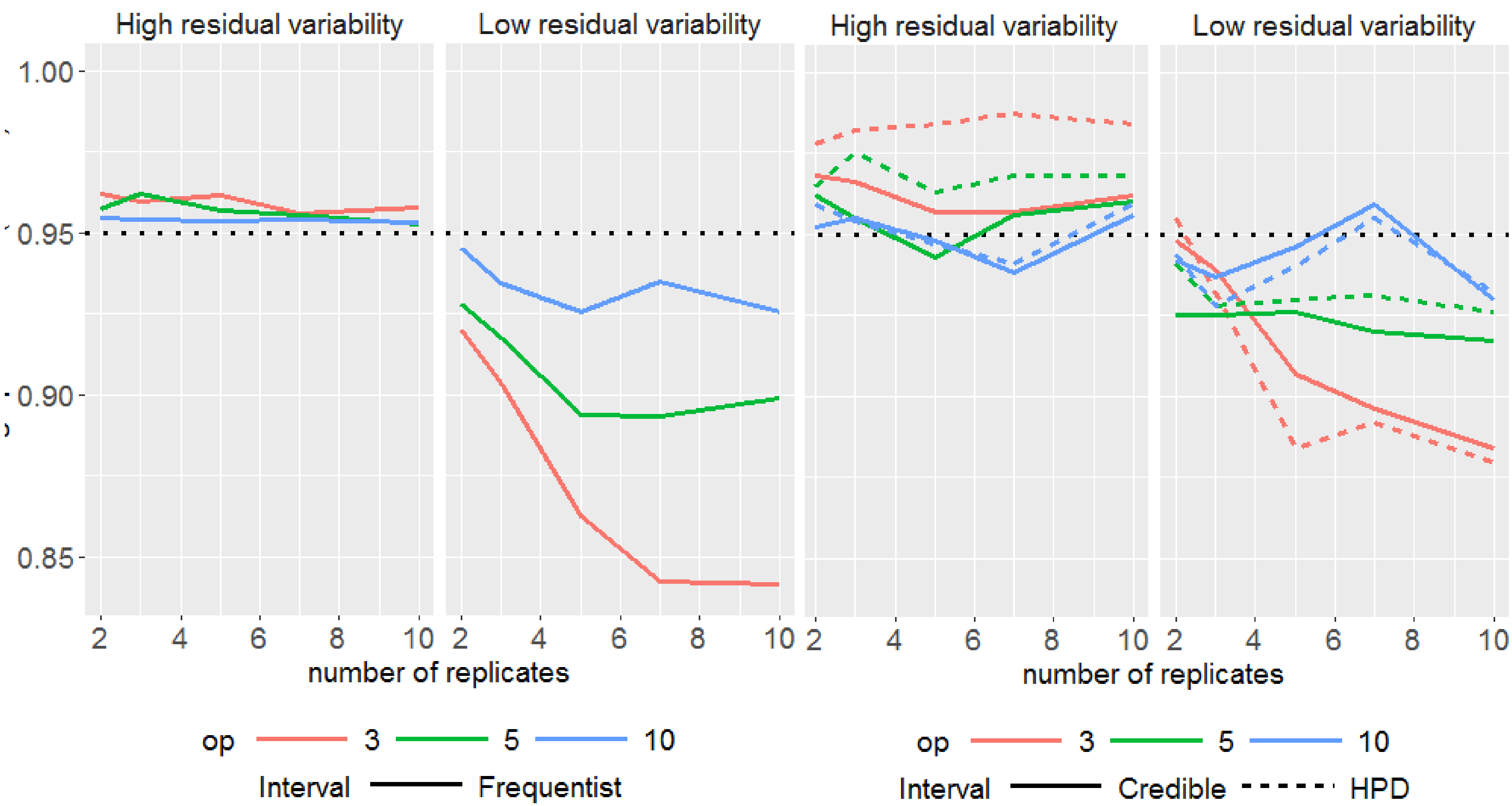
Reference level
= overall (unweighted) mean
→ JMP



Reference level
= a given level fixed effect
→ SAS, R,…

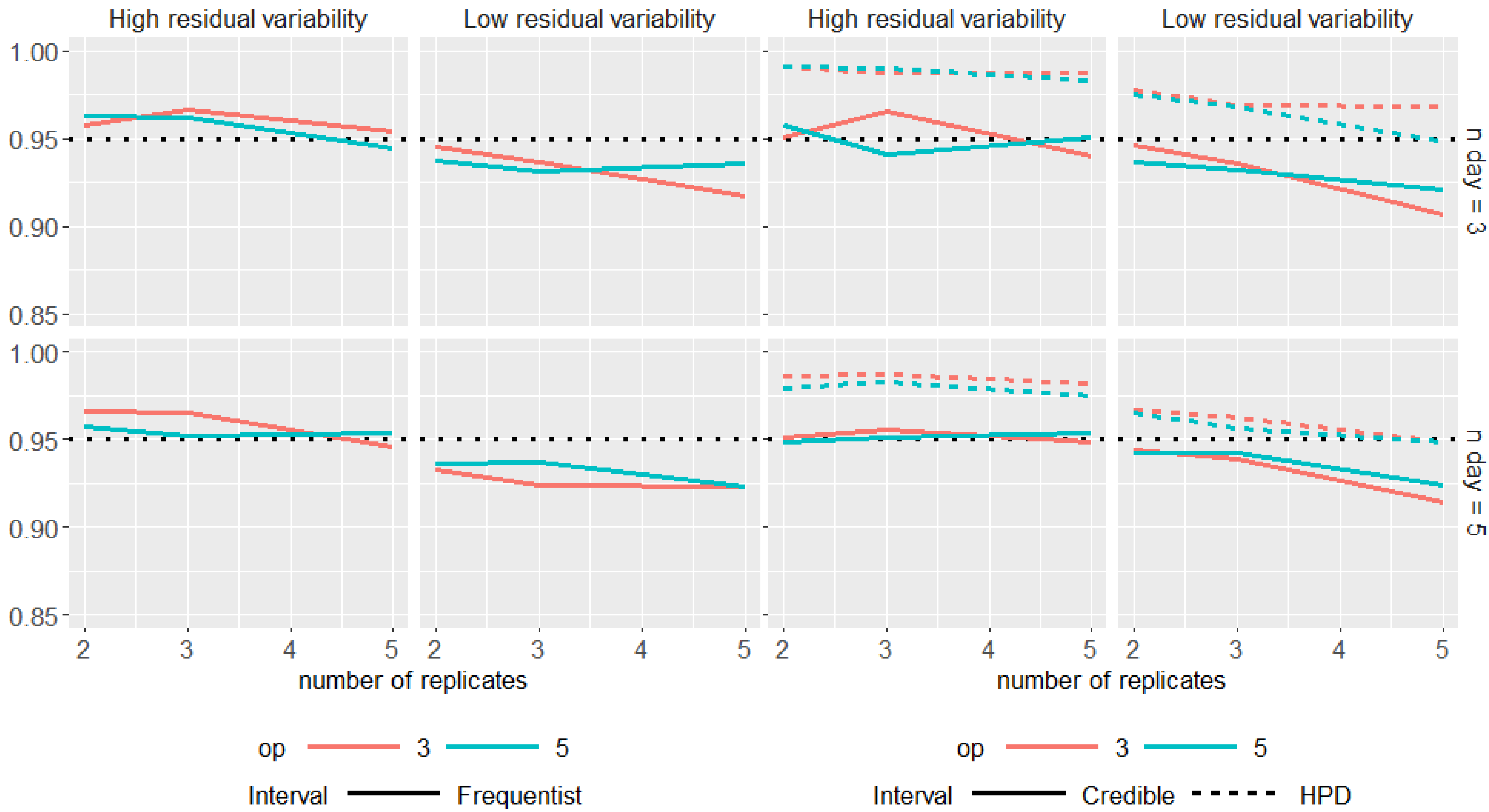


Total CV – Coverage probabilities (95%)
1 random variable



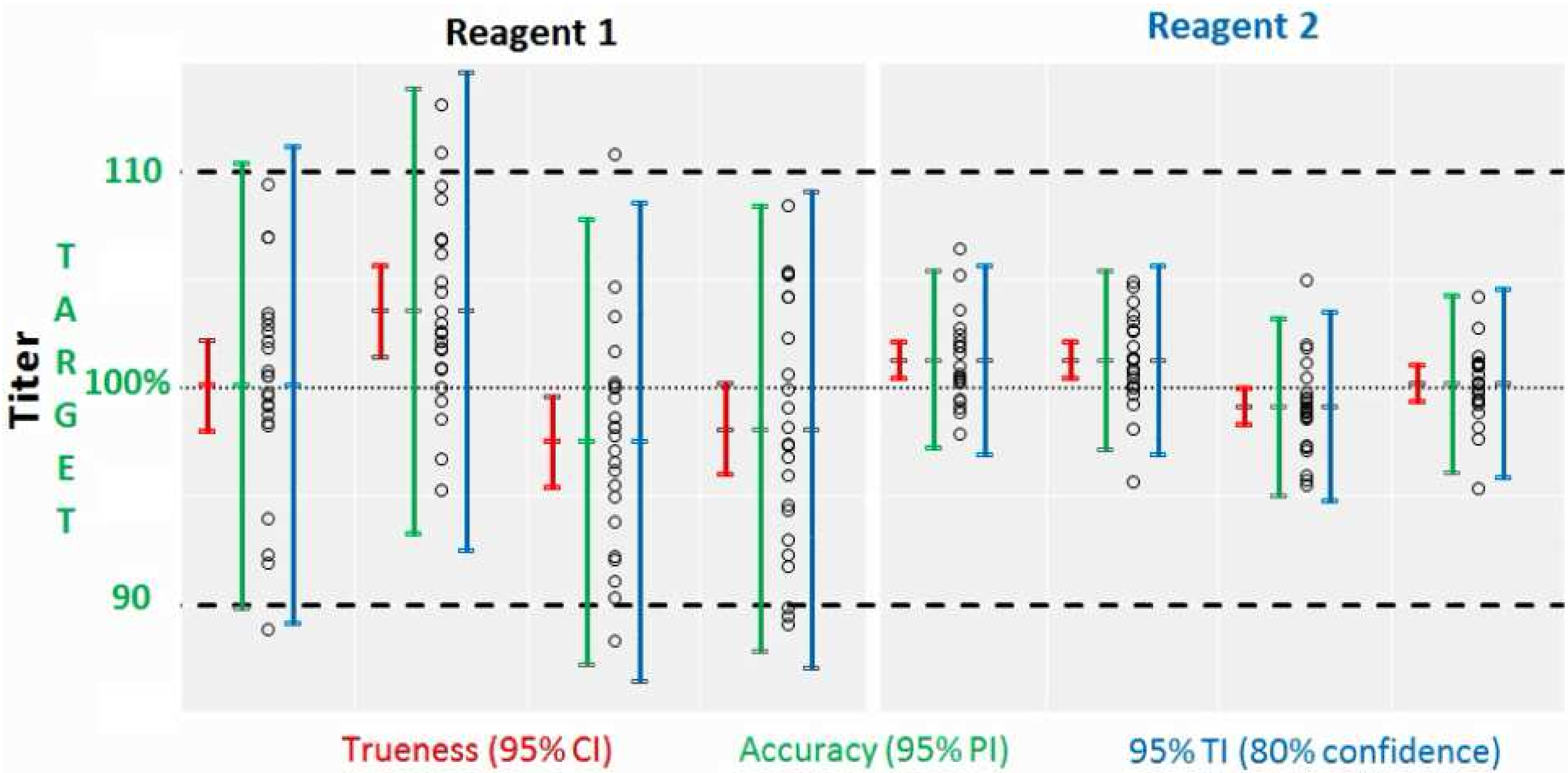
- Better coverage probabilities for high residual variability and high number of levels
- Bayesian better for low residual variability

Total CV – Coverage probabilities (95%)
2 crossed random variables

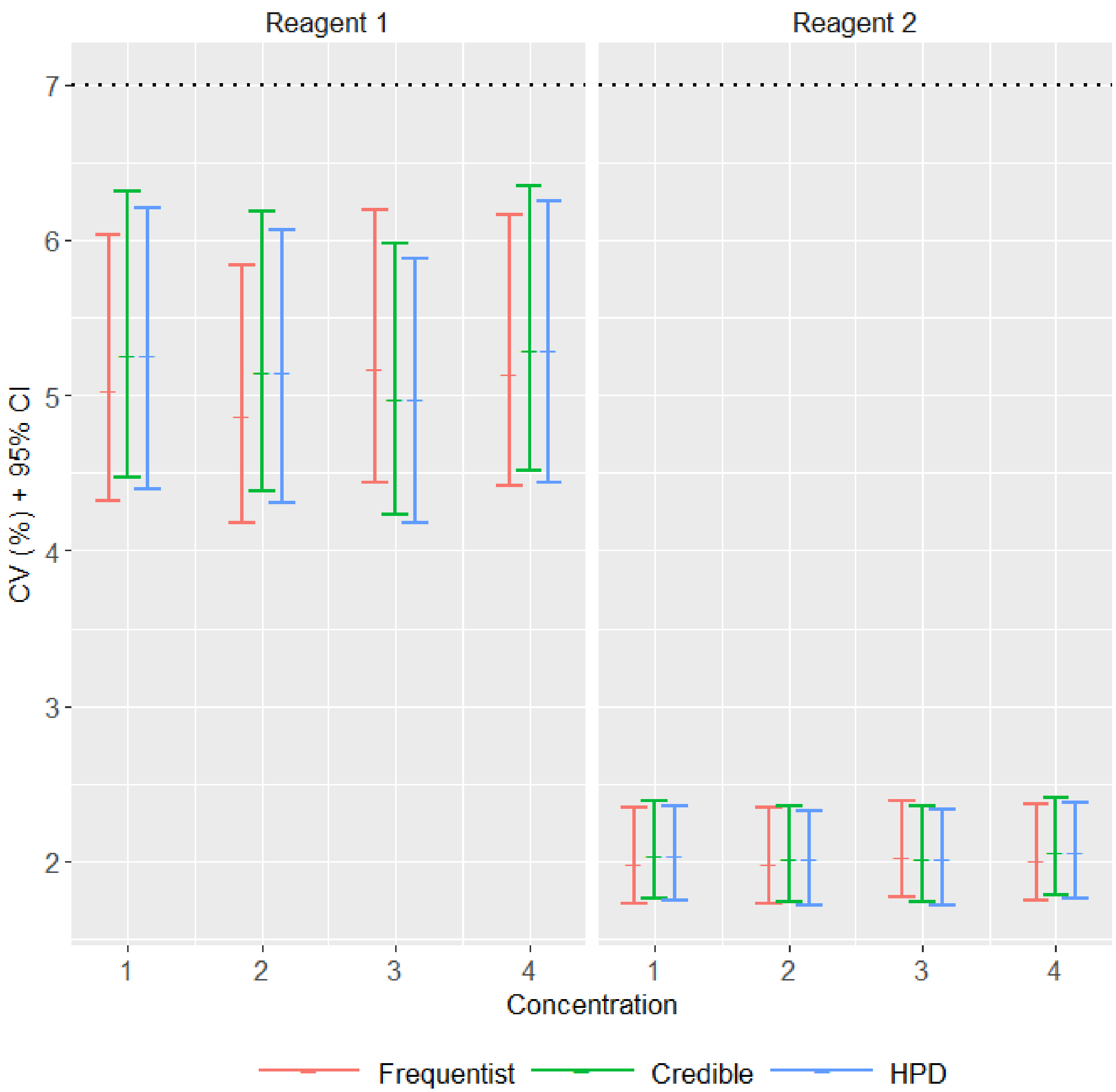


- Similar coverage probabilities Frequentist vs Bayesian
- Credible Intervals better than HPD

Assay Validation – Results Plot



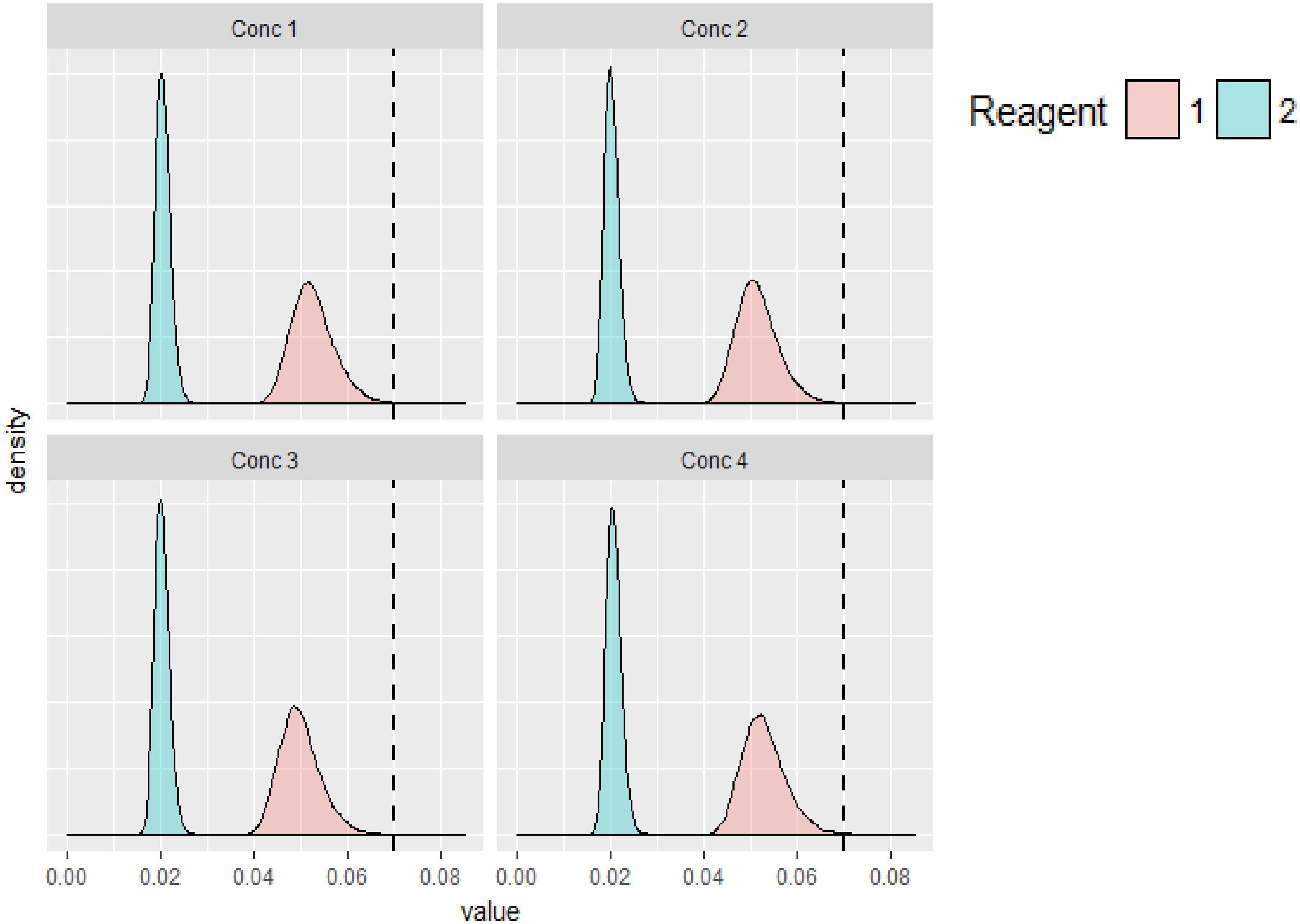
Assay Validation – Results CV - Plot



Credible & HPD
Bayesian intervals
(weakly priors)
are similar to the
frequentist intervals

Intermediate Precision is
higher for Reagent 1, but
all CVs (+95%CI) are
lower than 7%

Posterior Distribution – CV Intermediate Precision



Conclusions

Frequentist

- CI for CV is challenging
  - McKay formula adaptation
    - Analytical formula (direct)

- Calculate CV is straightforward
- Intervals obtained from posterior
- Weakly informative prior provides similar results to frequentist

Bayesian

Trueness and accuracy profile,
but also intermediate precision
are very useful in assay
qualification and validation

:-) :-) :-) :-) :-) :-) :-) :-) :-) :-) :-)
:-) *Give us your feedback* :-)
:-) :-) :-) :-) :-) :-) :-) :-) :-) :-) :-)

Last but not least

Reference

BG Francq, D Lin, W Hoyer. ***Confidence, Prediction and Tolerance in a General Linear Mixed Model.*** *Under review in Stat. in Med.*

Acknowledgement

GSK Technical R&D statisticians

Conflict of interest

This work was sponsored by GlaxoSmithKline Biologicals SA.
BGF, DL, MLB and WH are employees of the GSK group of companies.
MS is a student at UCLouvain University and was performing a traineeship at GSK.