6^{th} Open Immunogenicity Congress Lisbon 2014

Cutpoint estimation assuming a mixing distribution within the mixed model - using R

Ludwig A. Hothorn
Leibniz University Hannover, Germany
hothorn@biostat.uni-hannover.de
with F. Schaarschmidt (Han), Th. Jaki (Lanc), B. Gruen (Linz) and M. Hofmann (Basel)

February 20, 2014

Topics I

- Mixing distribution:
  naive samples may contain both ADA+ and ADA-
- Mixed model:
  multiple microtiter plates as random factor
- Challenge:
  combination of both approaches to estimate a prediction interval

The problem I

- First idea: Cut-point estimation is rather simple

  A screening cutpoint (SCP) is defined for a $N(\mu_i, \sigma^2)$ endpoint x in m historical samples, as an one-sided prediction interval for k of n future samples as:

$$SCP = uCP^{k,n}_{pred} = \overline{x} + s \times r_{k,n,m,1-\alpha/n}$$

  the quantile $r_{k,n,m,1-\alpha/n}$ can be estimated by the R-library `pred.intervals` (Hothorn et al. (2009))
  or even simpler $r = z_{1-\alpha} = 1.6445$

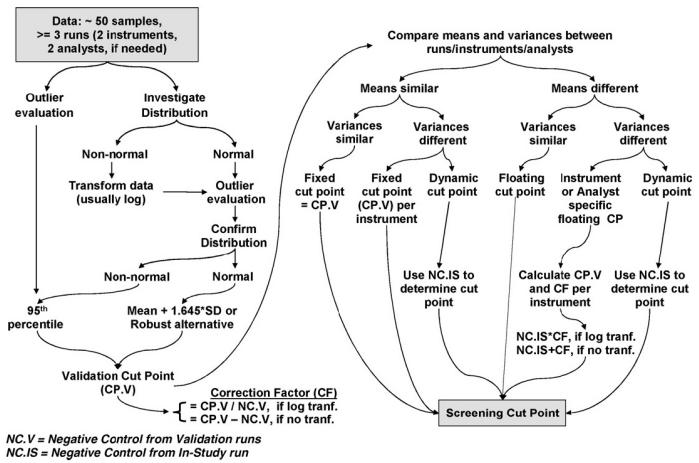
  ($uCP^{k,n}_{pred}$ ... upper prediction limit as cutpoint)

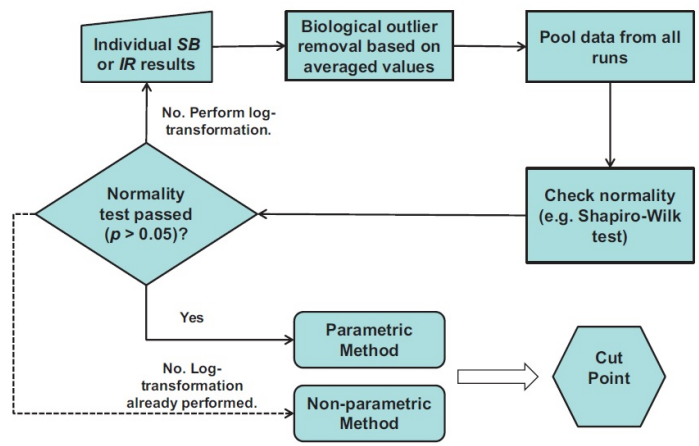
The problem II

- Second idea: Really simple?
  - m is rather small (e.g. $m = 50$)
    to achieve acceptable correct model selection rates for:
    1. *normal or not*
    2. *outlier or not (and which one)*
    3. *variance homogeneity or not*
    4. *homogeneity after normalization or not*

    in decision trees

    e.g. Shankar et al. (2008), Kubiak et al. (2013)





The problem III

- The assumption $N(\mu_i, \sigma^2)$ is questionable:
  log/normal?, homo/hetero variances?, unimodal/bimodal distibution? (ADA+ **and** ADA- in naive sample possible)
  ADA+ subjects shift SCP to larger values $\Rightarrow$ increase of f- rate
- The common design uses more than one plate:
  1. $>$ 96 samples needed
  2. possibly spiked and un-spiked on a plate for simultaneous specificity/confirmatory cut point determination
  3. possibly males and females on the same plate
  4. possibly positive controls on the same plate
  5. secondary factors to consider (prediction!): analysts, instruments, days,...
     AND: different hierarchies between these factors possible
- I.e. multiple assays with commonly different designs
- Normalization using NC is common- however this does not necessarily result in near to zero variance components, i.e. mixed model may be necessary
- Therefore SCP estimation may be complicated

The problem IV

- **Here we propose an approach:**
  - ⋆ Assuming a bimodal distribution. Selecting the ADA- samples for SCP estimation assuming a mixing distribution model
  - ⋆ SCP estimation in the mixed model, i.e. taking variability between runs (or analyst/plates, or ...) into account.
    May be to complicated for biologists:
    i) need for simplification, ii) proposing an appropriate design
  - ⋆ Modeling variance heterogeneity
  - ⋆ Instead of simple method above, even new statistical methods must be worked out, i.e.
    I) mixing distribution with random factors and heterogeneous variances,
    II) prediction intervals in mixed model

  Making available: using R

ADA assay as binary diagnostic test I

- Binary diagnostic tests are wide-spread used in medicine
  Similarity to ADA-assays:
  - classifying new samples into + or -
  - using a cutpoint for a continuous endpoint, so-called reference values
  - sometimes without gold standard (idea: considering multiple tests)
  - here even without $true^+$ and $true^-$ samples
  - therefore, explicit quantification of the error rates is not possible
  - but: **a smaller cutpoint results in a lower $false^-$ rate**
  - immunogenicity assays belong to safety assessment and therefore **controlling the $false^-$ is of primary importance**
  - therefore: smaller cutpoints should be preferred

Method I: Prediction interval in the mixed model I

- Room for confusion:
  the upper limits of all three type of intervals are defined **similar**
  $uSCP^{general} = \overline{x} + s \cdot quantile$ (see Hahn and Meeker (1991))
  - but a **Confidence interval** contains the population **mean** with a pre-specified confidence probability
  - but a **Tolerance interval** contains a specified proportion of future samples where the number of future samples needs not to be specified
  - and a **Prediction interval** for $k = 1$ of n future samples is appropriate (simplification for a single future observation is not necessary, but easy to harmonize)
- A **prediction interval** for a simple one-way layout with a naive variance estimate SD may be inappropriate
  - samples are splitted over plates (paired design)
  - even after normalization variance components between plates, analysts, days, devices,... may be not zero

Method I: Prediction interval in the mixed model II

- A) Prediction intervals for random effects models are needed
- Extensions of Hoffman and Berger (2011)
  For example:
  - $y_{ij} = \mu + a_i + b_j + ...$
  - $a_i \sim N(0, \sigma^2_{subject})$ variance between subjects, $i = 1, ..., n_{subject}$
  - $b_j \sim N(0, \sigma^2_{plate})$ variance between plates, $j = 1, ..., n_{plate}$
  - ... further variance components analyzed during method validation
- An upper limit that contains a single future observation (from the same population) with probability $(1 - \alpha)$:
- $\hat{\mu} + t_{1-\alpha, df_S} \sqrt{\hat{V}(y^*) + \hat{V}(\hat{\mu})}$ where
  - $\hat{V}(y^*)$ variance of a new observation y^*: the sum of variance components $\hat{V}(y^*) = \hat{\sigma}^2_{subject} + \hat{\sigma}^2_{plate} + ...$
  - $\hat{V}(\hat{\mu})$ variance of the estimated general mean, $\hat{\mu}$,
    $\hat{V}(\hat{\mu}) = \hat{\sigma}^2_{subject}/n_{subject} + \hat{\sigma}^2_{plate}/n_{plate} + ...$
  - the sum of both can be estimated as a weighted sum of the mean squares, MS, of an ANOVA table,

Method I: Prediction interval in the mixed model III

- the weights depend on the particular experimental design
- $t_{1-\alpha,df_S}$ is the $(1-\alpha)$ t-quantile with Satterthwaite (1941)-df

- For practical application, it is crucial:
  - to use standard experimental designs for ANOVA, with proper randomization
  - to correctly describe nesting or crossing of factors (subject, plate, ...) in the experiment
- R functions available for:
  - one factor
  - two factor hierarchical design
  - two factor crossed design with (and without) replications
  - three factors, with two crossed, third nested

Method I: Prediction interval in the mixed model IV

- B) Checking normality assumption in the mixed effects model
- Naive (realistic question): errors normal or log-normal distributed?

1. Likelihood ratio test for Box-Cox λ (Gurka et al. (2007))
   - ★ Scaled Box-Cox-transformation:
     $w_i = (y_i^{\lambda} - 1)/(\lambda \tilde{y}^{\lambda-1})$ if $\lambda \neq 0$
     $w_i = \tilde{y} \log(y_i)$ if $\lambda = 0$
   - ★ Notice: $\lambda = 1 \Rightarrow$ normal distribution, $\lambda = 0 \Rightarrow$ log-normal d.
   - ★ Fit models by maximum likelihood (not REML)
   - ★ Estimate the best $\hat{\lambda}$ (ML-estimator)
   - ★ Likelihood ratio tests:
   - ★ Test deviation from normality:
     $T = -2(L(w, \lambda = 1) - L(w, \lambda = \hat{\lambda})) \sim \chi^2_{df=1}$
   - ★ Test deviation from lognormality:
     $T = -2(L(w, \lambda = 0) - L(w, \lambda = \hat{\lambda})) \sim \chi^2_{df=1}$

Method I: Prediction interval in the mixed model V

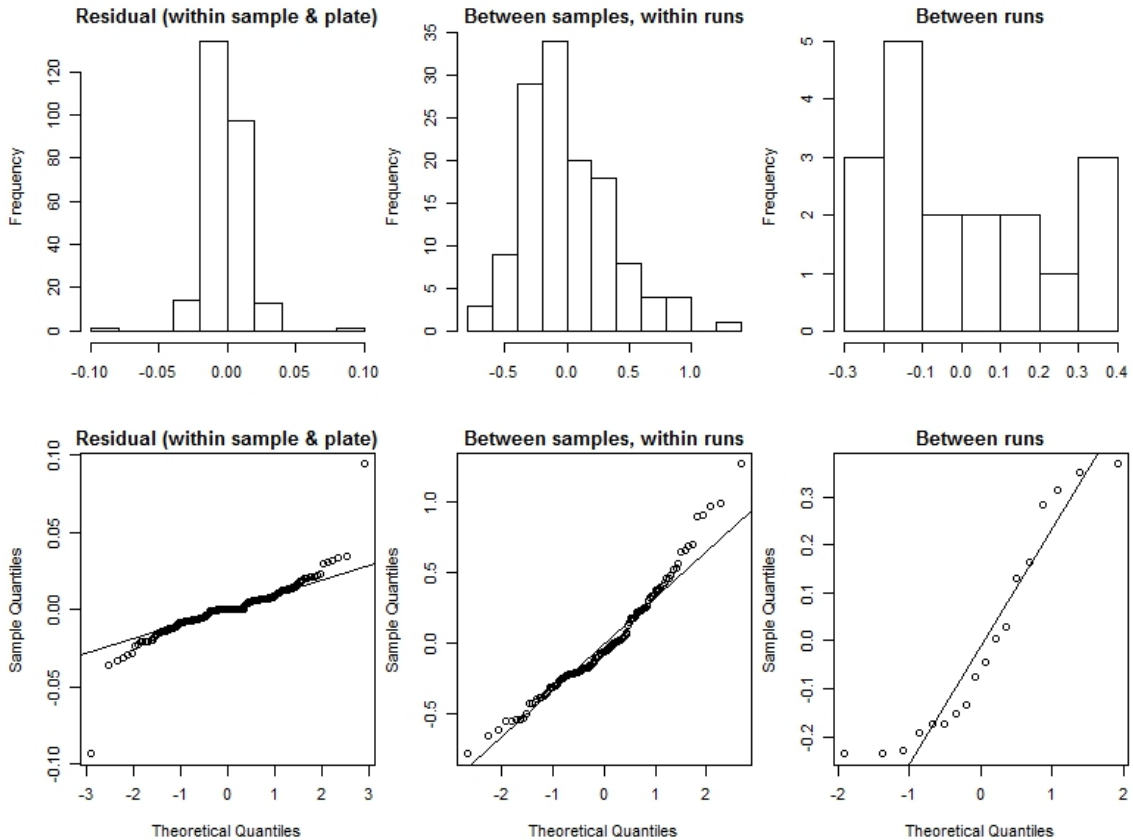
2. Visual assessment: Q-Q-plots and residual plots for random effects
3. An example: 18 runs; samples, considered nested in runs; techn. replicates within sample & run
4. estimated Box-Cox parameter: $\hat{\lambda} = -0.9$, i.e. even more skewed than lognormal

H_0	H_A	T_{LRT}	$\Pr(> \chi_{df=1})$
Normal ($\lambda = 1$)	Dev. from Normal	251.44	< 0.0001
Lognormal ($\lambda = 0$)	Dev. from Lognormal	52.05	< 0.0001

Log-transformation is the better choice, although not perfect

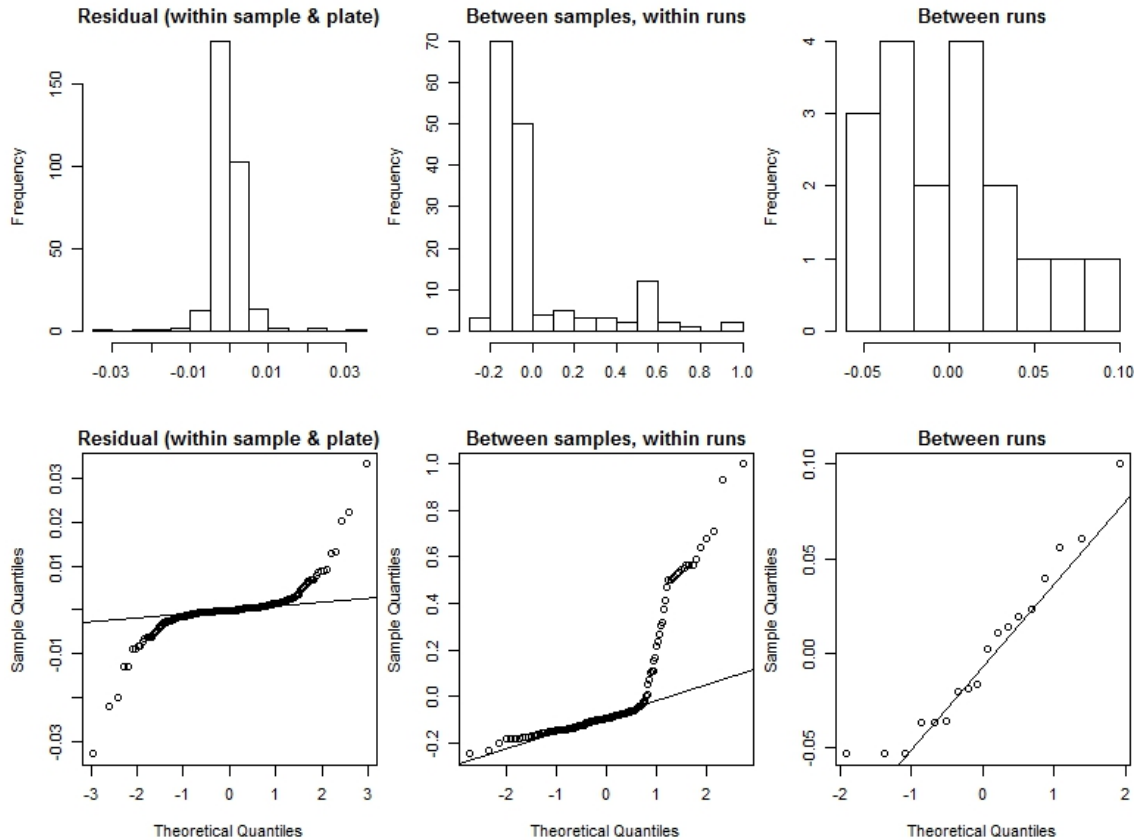
Method I: Prediction interval in the mixed model VI

5. Log-normal transformed data



Method I: Prediction interval in the mixed model VII

6. Original data (no transformation)



7. I.e. even to check whether normal or log-normal is not is simple job in the mixed model. Notice, the consequences on the cutpoint can be drastic

Method II: Normalization I

- Each plate contains NC samples
- Normalization against NC is common:

1. normal distribution $z_{ij} = y_{ij} - \bar{y}_{i,j=NC}$ (for i plates)
2. log-normal distribution $z_{ij} = w_{ij} - \bar{w}_{i,j=NC}, (w_{ij} = log(y_{ij}))$
3. possible: $z_{ij} = y_{ij}/\bar{y}_{i,j=NC}$ or $z_{ij} = y_{ij} - median(y_{i,j=NC})$ (not here)
4. lack-of-fit test (normal vs. log-normal): too low power for $m = 50$ (and complicated in hierarchical designs Xu et al. (2013))
5. therefore parallel estimation of SCP^{normal}, $SCP^{log-normal}$ and testing Box-Cox parameter (see above)

Method II: Normalization II

- After normalization variances between plates (analysts, devices, ...) may tend to zero or not Zhang et al. (2013)

		Analyst	Day	Plate	Residual	Total
ECL values	Variance	179.1	616.3	95.0	155.0	1045.5
	Percent	17%	59%	9%	15%	100%
Normalized ECL	Variance	0	0	0	0.0020	0.0010
	Percent	0	0	0	100%	100%

- But no powerful test on *no variances exists* in the mixed model (Wood (2013))
- I.e. we need $\Rightarrow$ prediction interval estimation in the mixed model: whether variance components tend to zero or not

Method III: Mixing distribution I

- Remember: too large values shift SCP up, which increase $false^-$ rate. This should be avoided
- Outlier tests commonly used
- Why outlier tests may be inappropriate:
  - outlier tests work perfect if the underlying distribution is known. Here even normal/log?
  - outlier tests should identify a tiny proportion of *extreme* values. Assays exist with $\hat{p}_{ADA+}$50%.
  - Even for 10% ADA+ taking the 95% percentile into account $\Rightarrow$ outlier test?
  - a repeated use of outlier tests to achieve an unimodal (normal) distribution if a mixing distribution exists seem to be questionable (Holland et al. (2013))
  - no simple one-way layout exists: it can be complicated hierarchically

Method III: Mixing distribution II

- Alternative: A) Mixing distribution approach Jaki et al. (2011). Allow mixing distribution between negative values and positive values under healthy volunteer samples already

$$Y = (1-p)Y_{ADA-} + pY_{ADA+}$$

- $Y_{negative} \sim F(\mu_{negative}, \sigma^2_{negative})$, $Y_{positive} \sim G(\mu_{positive}, \sigma^2_{positive})$ for some known distributions F and G
- The $(1-p)y_{negative}$ data are selected based on estimated model and using only these negative data a quantile approach is used for cut-point estimation
- This can be performed with the R-package `gamlss.mx` for an pseudo-one-way layout after pooling over plates
- restricted to one-way layout

Method III: Mixing distribution III

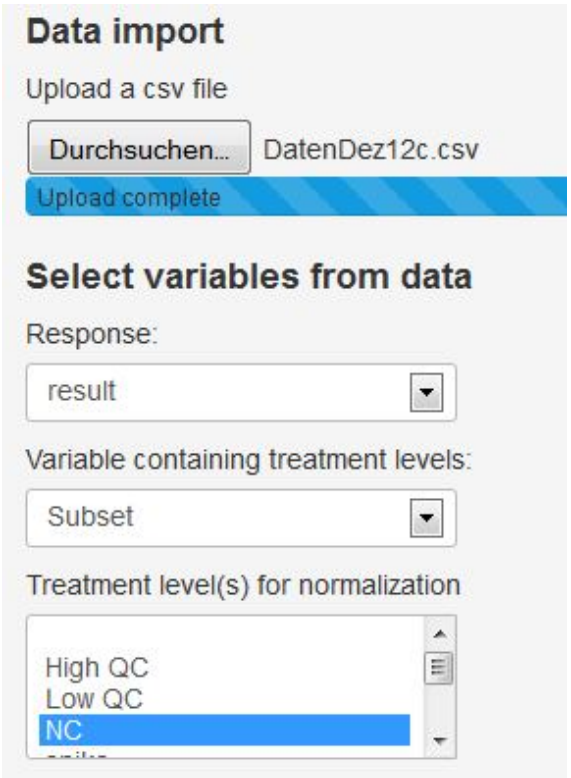
- Alternative: B) Mixing distribution approach in the mixed model Grun and Leisch (2007, 2008, 2009); Grun et al. (2012); Scharl et al. (2010)
- Considering heterogeneous variances:
  - between IDs, between replicated samples
  - Random effects in 2-component mixture model: both, equal random effects and equal residual variance
  - Equal random effects, different residual variance
  - Equal residual variance, different random effects
  - Both, different random effects and different residual variance
- Both approaches available as R programs
- Depending on the data condition and the particular design **transformation** $\Rightarrow$ **normalization** $\Rightarrow$ **selected ADA- population assuming ...** $\Rightarrow$ **estimation of the cutpoint in the mixed model** may be complicated

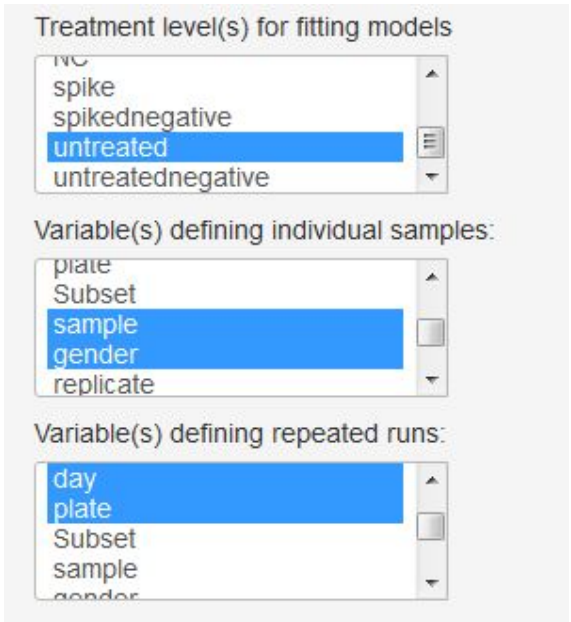
A user-friendly R program I

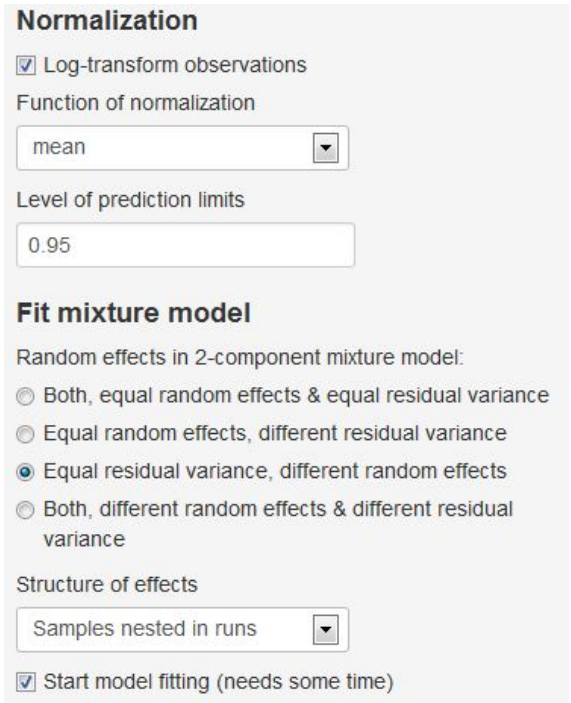
- Normalization assuming normal **and** log-normal distribution
- Testing normal vs. log-normal distribution in the mixed model
- Mixing distribution assuming:
  - bimodal distribution
  - heterogeneous variances
  - random factor(s)
  - selecting ADA^{-} samples for SCP estimation
- Prediction limit in the mixed model for nested or crossed between plate effects

A user-friendly R program II

- Interactive web application (shiny): data, variables and model options



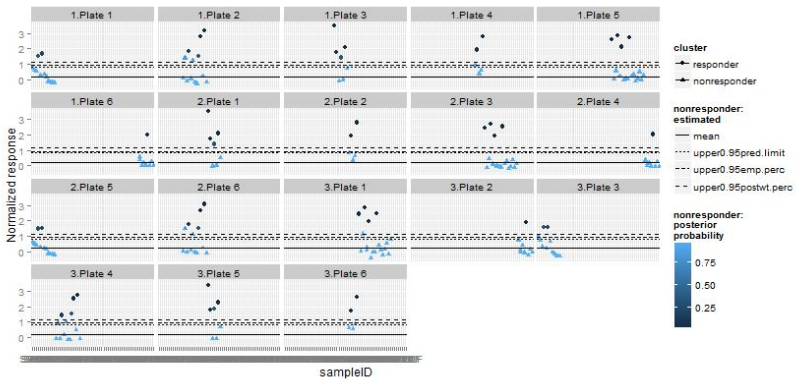




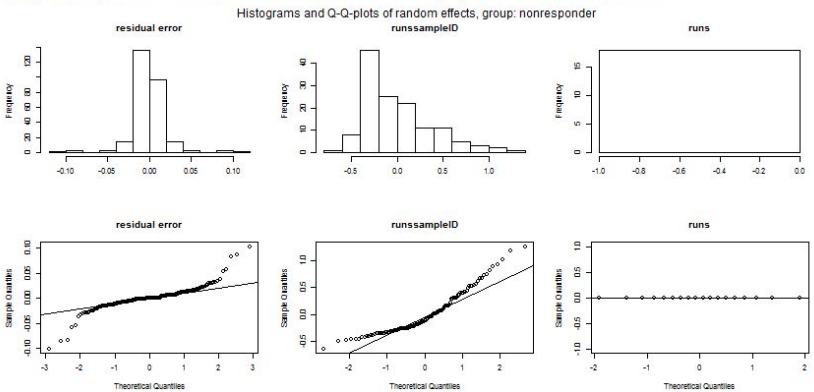
A user-friendly R program III

Two-component mixture model with random effects for sampleID and runs

Data used for fitting, prediction limits and posterior probability for subpopulation 'nonresponder'



Diagnostic plots for random effect estimates in subpopulation 'nonresponder'



Estimated mean, prediction limit and quantiles for 'nonresponder'

	value	group	estimated
1	0.20	nonresponder	mean
2	0.83	nonresponder	upper0.95pred.limit
3	1.13	nonresponder	upper0.95postwt.perc
4	0.93	nonresponder	upper0.95emp.perc

'pred.int': prediction limit (for 1 future observation) based on fitting a random effects model to those observations that were class as 'nonresponder' in the 2-component mixture model; 'postwt.perc' percentile of a sample the original observation, weighted by posterior probability to be member of group 'nonresponder'; 'emp.perc': percentile of those original observations that were class as 'nonresponder' in the 2-component mixture model.

Mixture model fit: parameter estimates and size of groups (a posteriori)

	labels	mean	V.ID	V.runs.in.ID	var.res	no.ID	no.obs
Comp.1	responder	2.21	0.34	0.02	0.00	17.00	102.00
Comp.2	nonresponder	0.38	0.19	0.02	0.00	45.00	270.00

Box-Cox-Lambda and LRT for normality and lognormality in mixed effects mixture model

	LogLikelihood	lambda
1	71.95	1.00
2	133.64	0.00
3	146.79	-0.80

3rd line: Box-Cox lambda, found by grid search from -3 to 3 by increment 0.1 for backtransformed response (after normalization classification)

	H0	HA	statLRT	Pr(> chi(df=1))
1	Normal (lambda=1)	Dev. from Normal	149.69	0.00
2	Lognormal (lambda=0)	Dev. from Lognormal	26.29	0.00

LRT for normality and lognormality in mixed model (Gurka, Edwards, Nylander-French, 2007)

Notice: non-responder ... ADA-

- R-code for more complex approach for statisticians: i) AIC-based model selection for different random factor formulation, ii) Different random factor formulations, iii) variance heterogeneities

Summary I

- Different assays with different designs and data conditions: we recommend a case-by-case analysis by a biostatistician instead of a simplified decision tree approach
- The decision makers are biologists and we understand their need for a simplified, robust approach. For some assays it works, for others not. Here, the danger of biased cutpoint estimation can be serious
- The heterogeneities in some assays are rather complex. Therefore, other approaches than SCP may be appropriate, e.g. for in-study data, classification, supervised or unsupervised learning
- A series of R programs and a web interface are available

Randomize samples within the plates

A nonparametrical prediction interval is available Frey (2013), but not in the mixed model

References I

Frey, J. (2013). Data-drivennonparametricpredictionintervals. *J. Stat.PLanning Inference*, 143:1039–1048.

Grun, B. and Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in r. *Computational Statistics & Data Analysis*, 51(11):5247–5252.

Grun, B. and Leisch, F. (2008). Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28(4):1–35.

Grun, B. and Leisch, F. (2009). Dealing with label switching in mixture models under genuine multimodality. *Journal of Multivariate Analysis*, 100(5):851–861.

Grun, B., Scharl, T., and Leisch, F. (2012). Modelling time course gene expression data with finite mixtures of linear additive models. *Bioinformatics*, 28(2):222–228.

Gurka, M. E., Edwards, L. J., and Nylander-French, L. (2007). Testing transformations in the linear mixed model. *Computational Statistics and Data Analysis*, 51:4297–4307.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals - A guide for practitioners*. John Wiley and Sons, Inc., New York.

Hoffman, D. and Berger, M. (2011). Statistical considerations for calculation of immunogenicity screening assay cut points. *Journal of Immunological Methods*, 373(1-2):200–208.

Holland, M. C., Wurthner, J. U., Morley, P. J., Birchler, M. A., Lambert, J., Albayaty, M., Serone, A. P., Wilson, R., Chen, Y., Forrest, R. M., Cordy, J. C., Lipson, D. A., and Bayliffe, A. I. (2013). Autoantibodies to variable heavy (v-h) chain ig sequences in humans impact the safety and clinical pharmacology of a v-h domain antibody antagonist of tnf-alpha receptor 1. *Journal of Clinical Immunology*, 33(7):1192–1203.

Hothorn, L. A., Gerhard, D., and Hofmann, M. (2009). Parametric and non-parametric prediction intervals based phase II control charts for repeated bioassay data. *BIOLOGICALS*, 37(5):323–330.

Jaki, T., Lawo, J.-P., Wolfsegger, M. J., Singer, J., Allacher, P., and Horling, F. (2011). A formal comparison of different methods for establishing cut points to distinguish positive and negative samples in immunoassays. *JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS*, 55(5):1148–1156.

Kubiak, R. J., Zhang, L. J., Zhang, J. C., Zhu, Y., Lee, N., Weichold, F. F., Yang, H., Abraham, V., Akufongwe, P. F., Hewitt, L., Robinson, S., Liu, W. Y., Liu, X., Patnaik, M. M., Spitz, S., Wu, Y. L., and Roskos, L. K. (2013). Correlation of screening and confirmatory results in tiered immunogenicity testing by solution-phase bridging assays. *Journal of Pharmaceutical and Biomedical Analysis*, 74:235–245.

References II

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, 6:309–316.

Scharl, T., Grun, B., and Leisch, F. (2010). Mixtures of regression models for time course gene expression data: evaluation of initialization and random effects. *Bioinformatics*, 26(3):370–377.

Shankar, G., Devanarayan, V., Amaravadi, L., Barrett, Y., Bowsher, R., Finco-Kent, D., Fiscella, M., Gorovits, B., Kirschner, S., Moxness, M., Parish, T., Quarmby, V., Smith, H., Smith, W., Zuckerman, L., and Koren, E. (2008). Recommendations for the validation of immunoassays used for detection of host antibodies against biotechnology products. *Journal of Pharmaceutical and Biomedical Analysis*, 48:1267–1281.

Wood, S. (2013). A simple test for random effects in regression models. *Biometrika*, 100.

Xu, W. L., Li, Y. W., and Song, D. W. (2013). Testing normality in mixed models using a transformation method. *Statistical Papers*, 54(1):71–84.

Zhang, L. J., Zhang, J. J., Kubiak, R. J., and Yang, H. (2013). Statistical methods and tool for cut point analysis in immunogenicity assays. *Journal of Immunological Methods*, 389(1-2):79–87.