Reports of the Institute of Biostatistics

No 01 / 2014

Leibniz Universität Hannover

Faculty of Natural Sciences

Titel: *Simple prediction intervals for one future observation for simple, balanced random effects models*

Authors: *F. Schaarschmidt, L.A. Hothorn*

1 Introduction

General purpose: define upper limits of measurements for a 'non-responder' population based on assays for ADA ('cutpoint') Experimental background: laboratory experiments are run that measure ADA with some assay system, where ADA-positive samples ('responders') tend to high measurements and ADA-negative samples ('non-responders') tend to low values in the laboratory experiments, there are two types structures that need to be addressed in the analysis: 1) there are biological samples (representing sera of different laboratory animals, or of different patients, etc.) and 2) structures related to the assay system, e.g. assay run on different days, by different analysts, with different plates, etc. Problem: given that the biological sample consists of 'non-responders' only (or, that a population of 'non-responders' has been identified within the sample under investigation), the estimation of a cutpoint should take experimental structures (biological samples, time, analysts, devices, plates, etc.) into account.

Following Hoffman and Berger (2011), we define the 'cutpoint' as an upper prediction limit that contains a single future observation with probability $(1-\alpha)$. The computation of such intervals is described for simple experimental layouts Hahn and Meeker (1991), and some special cases Satterthwaite (1941); Hoffman and Berger (2011). The purpose of this report is to formally describe these methods for a set of special cases, which may be relevant for experimental designs in laboratory experiments concerning ADA.

A major problem of available example data sets is that the experimental designs were unclear or fundamentally ill-designed, such that variance components of interest were confounded, or, at least, standard ANOVA models were not applicable. As a result, the set of necessary experimental layouts and corresponding models of potential interest is still unclear. The experimental layout and model underlying the methods in the initial reference Hoffman and Berger (2011) and all experimental layouts and models discussed in the following are considerably more simple than what might have lead to the real data that have been available so far.

In the following, along with a formal definition of statistical models, examples of corresponding experimental designs are described.

2 Models and corresponding experimental layout

2.1 One-way layout (h1)

$$\begin{array}{lcl} y_{ij} & = & \mu + a_i + e_{j(i)} \\ a_i & \sim & N(0, \sigma_a^2), i = 1, ..., I \\ e_{j(i)} & \sim & N(0, \sigma_e^2), j(i) = 1(i), ..., J(i) \end{array} \tag{1}$$

This model may be adequate for analyzing the following experimental design:

- a random sample of IJ patients is obtained and is randomly split into I subsets, $i = 1, ..., I$, each of size J,
- each subset i is assigned to one of I plates; within plate i, the jth patients sample is identified by $j(i)$,

i.e., patient j is NOT repeatedly analyzed at more than one plate, but each patient is analyzed only once at only one plate. Thus σ_a^2 is the between plate variance, while σ_e^2 contains the between patient variance as well as other variances within plate (e.g., spatial effects on the plate). This model with the application drafted above will not be able to cover the multitude of potential environmental effects. Thus, it has low practical relevance. However, it might still be used if I is chosen large, and the I levels are not just a number of plates, but a representative random sample of 'experimental runs' taken to cover all relevant environmental conditions (e.g. a random sample covering different labs, analysts, machines/devices/plates, days/times, climatic conditions etc.), given that the normality assumption is reasonable for such a combination of several effects. Moreover, this models is included because it has been used in basic references concerning the topic of prediction intervals.

2.2 Two-way hierarchical layout (h2)

$$\begin{array}{lcl} y_{ijk} & = & \mu + a_i + b_{j(i)} + e_{k(ij)} \\ a_i & \sim & N(0, \sigma_a^2), i = 1, ..., I \\ b_{j(i)} & \sim & N(0, \sigma_b^2), j(i) = 1(i), ..., J(i) \\ e_{k(ij)} & \sim & N(0, \sigma_e^2), k(ij) = 1(ij), ..., K(ij) \end{array} \tag{2}$$

This model may be adequate for analyzing the following experimental design:

- a random sample of IJ patients is obtained and is randomly split into I subsets, $i = 1, ..., I$, each of size J,
- each subset i is assigned to one of I plates; within plate i the jth patients sample is identified by $j(i)$,
- each patient $j(i)$ is repeatedly analyzed K times, $k = 1, ..., K$, but only at the assigned plate i, not across different plates.

Details for estimation of μ, and variance components in an ANOVA setting can be found in (Sahai and Ageel, 2000, p.352, Tab. 6.2; p.358).

2.3 Balanced two-way completely cross-classified, no replication per cell (c1)

$$\begin{array}{lcl} y_{ij} & = & \mu + a_i + b_j + e_{ij} \\ a_i & \sim & N(0, \sigma_a^2), i = 1, ..., I \\ b_j & \sim & N(0, \sigma_b^2), j = 1, ..., J \\ e_{ij} & \sim & N(0, \sigma_e^2) \end{array} \tag{3}$$

This is the model considered in Hoffman and Berger (2011), details for ANOVA, mean squares and variance of the general mean can be found in (Sahai and Ageel, 2000, p.134, Tab. 3.2; p.140-141). It may by adequate for analyzing the following experimental design:

- a random sample of I patients is obtained
- analysis is performed by (a random sample of) J operators
- each patient $i = 1, ..., I$, is analyzed by each operator $j = 1, ..., J$ exactly once, i.e. one observation y_{ij} is available for each patient from each operator

Of course, patients $i = 1, ..., I$ might be other random samples from the populations of interest, and what is called operator $j = 1, ..., J$ might be plates, days, analytical devices or whatever. Samples $i = 1, ..., I$ should be analyzed in a new random sequence over time (allocated in a new spatial randomization on plates) within each operator/plate/day $j = 1, ..., J$, such that errors associated with time course or spatial effects on plates are analyzed in σ_e^2; otherwise, the between patient variance σ_a^2 may be confounded with any error structure associated with time within day, learning effects within operators, or spatial effects on the plates.

2.4 Two-way completely cross-classified, with replication per combination (c2)

$$\begin{array}{lcl} y_{ijk} & = & \mu + a_i + b_j + (ab)_{(ij)} + e_{ijk} \\ a_i & \sim & N(0, \sigma_a^2), i = 1, ..., I \\ b_j & \sim & N(0, \sigma_b^2), j = 1, ..., J \\ (ab)_{ij} & \sim & N(0, \sigma_{ab}^2) \\ e_{ijk} & \sim & N(0, \sigma_e^2) \end{array} \tag{4}$$

This model may be adequate for analyzing the following experimental design:

- a random sample of I patients is obtained
- analysis is performed by J operators
- each patient $i = 1, ..., I$ is analyzed by each operator $j = 1, ..., J$ exactly K times

where the sequence of the individual observations ijk is randomized over time or with respect to the spatial allocation on plates/devices. Again, if the sequence of patients is not randomized within operator j (but all operators analyze the patients in the same sequence), the estimated between patient variability σ_a^2 will be confounded by consistent temporal effects (due to effects associated with time, learning) within operators. Similarly, b_j are J plates, not J operators: if the spatial allocation of patients on the plates is the same on all plates, the between patient variance estimate σ_a^2 will be confounded by consistent spatial effects on the plates. If the K replications per patient i are not randomized over the possible positions on plates j, but are all allocated next to each other on a given plate, spatial effects on the plate will confound the estimated variance component σ_{ab}^2. For details of estimating mean and variance components in this design, see (Sahai and Ageel, 2000, p.194, Tab. 4.2; p.208, Eq.(4.8.5)).

2.5 Three-way layout with two factors crossed, and one nested (ch3, Shankar)

- samples of K patients are split and each patient is analyzed by each of I analysts (such that patient and analyst are cross-classified, i.e., it is reasonable to assume that systematic differences between the analyst effect all patients additively)
- each analyst analyses each patient at J different plates, such that patient is also crossed with plate, but plate is nested within analyst (i.e. it is reasonable to assume that systematic differences between plate carries to all patients, whereas it is not reasonable to assume that plate $j = 1$ in analyst $i = 1$ shares a common effect with plate $j = 1$ in analyst $i = 2$),
- finally, for each patient and each plate (within analyst), L technical replications are obtained

$$
\begin{array}{lcl}
y_{ijkl} & = & \mu + a_i + b_{j(i)} + c_k + ac_{ik} + bc_{jk(i)} + e_{l(ijk)} \\
a_i & \sim & N(0, \sigma_a^2), i = 1, ..., I \\
b_{j(i)} & \sim & N(0, \sigma_{b(a)}^2), j = 1(i), ..., J(i) \\
d_k & \sim & N(0, \sigma_d^2), k = 1, ..., K \\
ad_{ik} & \sim & N(0, \sigma_{ad}^2) \\
bd_{jk(i)} & \sim & N(0, \sigma_{bd(a)}^2) \\
e_{l(ijk)} & \sim & N(0, \sigma_e^2), l(ijk) = 1(ijk), ..., L(ijk)
\end{array} \tag{5}
$$

I.e., σ_a^2 is the overall between-analyst variance, $\sigma_{b(a)}^2$ is the between-plate variance (i.e., mean deviation of plates w.r.t. each analysts overall mean), σ_d^2 is the overall between-patient variance. σ_{ad}^2 is the variance due to lack of additivity between overall analyst and overall patients effects (i.e, due to 'inconsistency' of patients effects between analysts), $\sigma_{bd(a)}^2$ is the variance due to inconsistency of patients effects between plates within a given analyst. The residual error (the deviation of technical replicates from a given mean of patient k in plate j of analyst i) is estimated by σ_e^2. For the details for sum of squares, expected mean squares (allowing to derive estimates for the variance components and variance of the grand mean etsimator) see (Sahai and Ageel, 2000, p.431-435, Tab. 8.1, p.436).

3 Prediction intervals for a single future observation

Based on mean squares of a classical analysis of variance, (e.g. Satterthwaite, 1941; Hoffman and Berger, 2011) describe prediction intervals. Hoffman and Berger (2011) do not give a reference or an analysis/discussion of the small sample performance of this method. Hoffman and Berger (2011) do not state what exactly this interval is supposed to cover (a single (M=1) future observation, mean of a M future observations, all of M future observations (compare Hahn and Meeker, 1991, p.61-63)).

This section recalls simple methods to construct prediction intervals that are supposed to cover a single future observation, y^*, with probability $(1 - \alpha)$. They have the general form

$$
\hat{\mu} \pm t_{1-\alpha/2, df_S} \sqrt{\hat{V}(y^*) + \hat{V}(\hat{\mu})} \tag{6}
$$

where

- $\hat{\mu}$ is the estimate for the general mean μ,
- $t_{1-\alpha, df_S}$ is the $(1-\alpha)$ quantile of a t-distribution with Satterthwaite degree of freedom, df_S, that is to be specified below,
- $\hat{V}(y^*)$ is estimated variance of a new observation y^* (from the same population) with respect to μ, usually the sum of variance components in the model of interest, and
- $\hat{V}(\hat{\mu})$ is the estimated variance of the estimated general mean, $\hat{\mu}$, with respect to the true mean μ; it is usually the sum of variance components, weighted by the corresponding sample size components of the validation experiment.

For example, based on model c1 (Eq.3) considered by Hoffman and Berger (2011), the variance term of the prediction intervals is

$$\hat{V}(y^*) + \hat{V}(\hat{\mu}) = \left[\hat{\sigma}_a^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2\right] + \left[\hat{\sigma}_a^2/I + \hat{\sigma}_b^2/J + \hat{\sigma}_e^2/(IJ)\right]$$

$$\hat{V}(y^*) + \hat{V}(\hat{\mu}) = \hat{\sigma}_a^2(1 + 1/I) + \hat{\sigma}_b^2(1 + 1/J) + \hat{\sigma}_e^2(1 + 1/(IJ)).$$

Obviously,

- as the sample size components I and J increase, the contribution due to the variance of the general mean, $\hat{V}(\hat{\mu})$, decreases. Then, a $(1-\alpha)$ upper prediction limit will become close the $(1-\alpha)$ quantile of the estimated normal density, $N(\hat{\mu}, \hat{\sigma}^2 = (\hat{\sigma}_a^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2))$; particularly, if the dominating variance components are associated with large corresponding sample size components.

- the variance term is weighted sum of variance components, where the single variance components are associated with different degrees of freedom, suggesting the Satterthwaite approximation of the overall df. The weights of the individual components approach 1 if the sample size components, I, J increase.

3.1 Classical analysis of variance

Unfortunately, in an ANOVA setting, only $\hat{\sigma}_e^2$ is directly estimated via the mean square error, MS_E. The estimators of the remaining variance components are already weighted linear combinations of the mean squares from the analysis of variances. Satterthwaite (1941); Hoffman and Berger (2011) derive the approximated degree of freedom directly from the mean squares, not from the estimators of the variance components. Following this approach for the models above leads to the following estimators:

where $c = 1, ..., C$ is the index of variance components in a given model. Note that MS^* denotes mean squares without the constant sample size weights they normally contain in ANOVA tables, i.e., those MS marked with * are not directly written in the usual ANOVA tables.

Using these mean squares and the corresponding weights, the variance term of the prediction interval is

$$(\hat{V}(y^*) + \hat{V}(\hat{\mu})) = \sum_{c=1}^{C} w_c MS_c \tag{7}$$

and the corresponding Satterthwaite approximation of the degree of freedom is

$$df_S = \frac{\left(\sum_{c=1}^{C} w_c MS_c\right)^2}{\sum_{c=1}^{C} \frac{(w_c MS_c)^2}{df_c}}, \tag{8}$$

Equation (6) with Model h1 (Eq. 1) and the corresponding mean squares and weights is already given in (Satterthwaite, 1941, p.314-315). It is a generalization of the prediction interval supposed to contain a single future observation with probability $(1 - \alpha)$ as provided by (Hahn and Meeker, 1991, p.61, Eq.(4.2)) given a single, unstructured random sample.

3.2 Using estimates from a mixed model fit

Assume that the models above are fitted in a mixed model and estimates for the general mean, $\hat{\mu}$, and the variance components $\hat{\sigma}_c^2$ are available. One may obtain the same variance term for the prediction interval by a weighted linear combination of the variance components estimates,

$$\hat{V}(y^*) + \hat{V}(\hat{\mu}) = \sum_{c=1}^{C} w_c \hat{\sigma}_c^2, \tag{9}$$

using the weights in Table 1. Using a similar equation to obtain degrees of freedom,

$$df_S = \frac{\left(\sum_{c=1}^{C} w_c \hat{\sigma}_c^2\right)^2}{\sum_{c=1}^{C} \frac{(w_c \hat{\sigma}_c^2)^2}{df_c}}, \tag{10}$$

results in slightly different weighting of the components, and hence in a slightly different degree of freedom, df_S. Thus, using this approach leads to different approximate degree of freedom as compared to Satterthwaite (1941) for model h1 and Hoffman and Berger (2011) for model c1.

Model	MS_c	df_c	w_c
h1	$MS_E = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2}{IJ - I}$	$IJ - I$	$(1 - 1/J)$
	$MS^*_A = \frac{\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2}{I - 1}$	$I - 1$	$(1 + 1/I)$
h2	$MS_E = \frac{\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2}{IJK - IJ}$	$IJK - IJ$	$(1 - 1/K)$
	$MS^*_{B(A)} = \frac{\sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..})^2}{IJ - I}$	$IJ - I$	$(1 - 1/J)$
	$MS^*_A = \frac{\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2}{I - 1}$	$I - 1$	$(1 + 1/I)$
c1	$MS_E = \frac{\sum_i \sum_j (y_{ij} - (\bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}))^2}{(I-1)(J-1)}$	$(I-1)(J-1)$	$(1 - 1/I - 1/J - 1/(IJ))$
	$MS^*_B = \frac{\sum_j (\bar{y}_{j.} - \bar{y}_{..})^2}{J - 1}$	$J - 1$	$(1 + 1/J)$
	$MS^*_A = \frac{\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2}{I - 1}$	$I - 1$	$(1 + 1/I)$
c2	$MS_E = \frac{\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2}{IJ(K-1)}$	$IJ(K-1)$	$(1 - 1/K)$
	$MS^*_{AB} = \frac{\sum_i \sum_j (\bar{y}_{ij.} - (\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}))^2}{(I-1)(J-1)}$	$(I-1)(J-1)$	$(1 - 1/I - 1/J - 1/(IJ))$
	$MS^*_B = \frac{\sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2}{J - 1}$	$J - 1$	$(1 + 1/J)$
	$MS^*_A = \frac{\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2}{I - 1}$	$I - 1$	$(1 + 1/I)$
c2h1	$MS_E = \frac{\sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \bar{y}_{ijk.})^2}{IJK(L-1)}$	$IJK(L-1)$	$(1 - 1/L)$
Shankar	$MS^*_{BD(A)} = \frac{\sum_i \sum_j \sum_k (\bar{y}_{ijk.} - (\bar{y}_{ij..} + \bar{y}_{i.k.} - \bar{y}_{i...}))^2}{I(J-1)(K-1)}$	$I(J-1)(K-1)$	$(1 - 1/J - 1/K + 1/(JK))$
	$MS^*_{AD} = \frac{\sum_i \sum_k (\bar{y}_{i.k.} - (\bar{y}_{i...} + \bar{y}_{..k.} - \bar{y}_{....}))^2}{(I-1)(K-1)}$	$(I-1)(K-1)$	$(1 - 1/I - 1/K - 1/(IK))$
	$MS^*_{B(A)} = \frac{\sum_i \sum_j (\bar{y}_{ij..} - \bar{y}_{i...})^2}{I(J-1)}$	$I(J-1)$	$(1 - 1/(IJ))$
	$MS^*_D = \frac{\sum_k (\bar{y}_{..k.} - \bar{y}_{....})^2}{K - 1}$	$K - 1$	$(1 + 1/K)$
	$MS^*_A = \frac{\sum_i (\bar{y}_{i...} - \bar{y}_{....})^2}{I - 1}$	$I - 1$	$(1 + 1/I)$

Table 1: Variance components, degrees of freedom, and weights for the linear combination of variance components in Eq. 9

Model	σ_c^2	df_c	w_c
Model h1,	σ_a^2	$I-1$	$1+1/I$
Eq.(1)	σ_b^2	$IJ-I$	$1+1/IJ$
Model h2,	σ_a^2	$I-1$	$1+1/I$
Eq.(2)	σ_b^2	$IJ-I$	$1+1/IJ$
	σ_e^2	$IJK-IJ$	$1+1/IJK$
Model c1,	σ_a^2	$I-1$	$1+1/I$
Eq.(3)	σ_b^2	$J-1$	$1+1/J$
	σ_e^2	$(I-1)(J-1)$	$1+1/IJ$
Model c2 ,	σ_a^2	$I-1$	$1+1/I$
Eq.(4)	σ_b^2	$J-1$	$1+1/J$
	σ_{ab}^2	$(I-1)(J-1)$	$1+1/IJ$
	σ_e^2	$IJK-IJ$	$1+1/IJK$
Model c2h1 ,	σ_a^2	$I-1$	$1+1/I$
Eq.(4)	$\sigma_{b(a)}^2$	$I(J-1)$	$1+1/(IJ)$
	σ_d^2	$K-1$	$1+1/K$
	σ_{ad}^2	$(I-1)(K-1)$	$1+1/(IK)$
	$\sigma_{bd(a)}^2$	$I(J-1)(K-1)$	$1+1/(IJK)$
	σ_e^2	$IJK(L-1)$	$1+1/(IJKL)$

4 Validation of software

To roughly check the correctness of underlying R code and to compare the different types of degrees of freedom (based on estimates of variance components from mixed model fits (lmer) , and the published way based on ANOVA mean squares) a very limited simulation study has been run:

- For each model (h1, h2, c1, c2, Shankar), the number of replications (a, b, d, n) and the corresponding variance components (σ_a^2,..., σ_e^2) have been fixed at different values
- given a single model and parameter setting:
  - draw realizations of all the random components of a given model with the given number of replications, leading to a vector of observations $y_{...}$
  - compute an upper 0.95 prediction limit as described above: $y^u = \hat{\mu} + t_{0.95,df_S} \sqrt{\hat{V}(y^*) + \hat{V}(\hat{\mu})}$
  - draw a new, single realization of all random components of a given model, leading to a single new obseravtion y^*
  - check whether the new observation is in the prediction interval, $y^* \leq y^u$
- repeat this 1000 times, leading to an estimate for the probability that a single future observation is contained in the prediction interval

4.1 Results

For the few considered settings, the observed proportion of cases, where a future single observation is below the upper 0.95 prediction limit, varies around the pre-specified value 0.95, and rarely falls outside the rejection region of a two-tailed binomial test with $n = 1000$ for $H_0 : proportion = 0.95$.

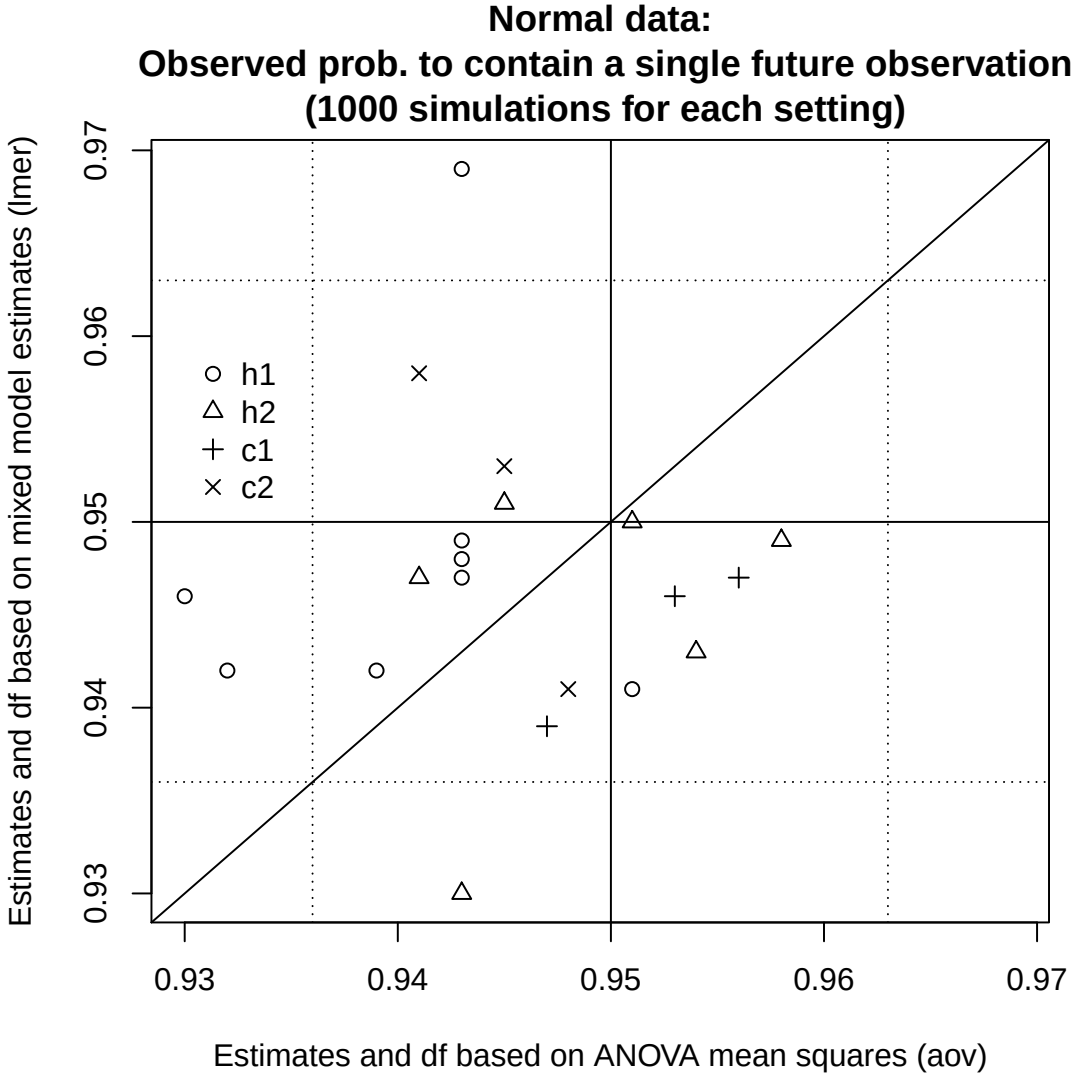

Figure 1: Observed proportion of cases where a single future observation is covered by a upper 0.95 prediction limit, for normal distributed data from models h1, h2, c1, c2. Proportions estimated from 1000 simulation runs for each model and parameters setting, dotted lines circumscribe the range of observed proportions for which $H_0 : proportion \neq 0.95$ can not be rejected in a two-tailed test at level $\alpha 0.05$, i.e. if a method would be exact, 95% of the simulation results fall between the two dotted lines.

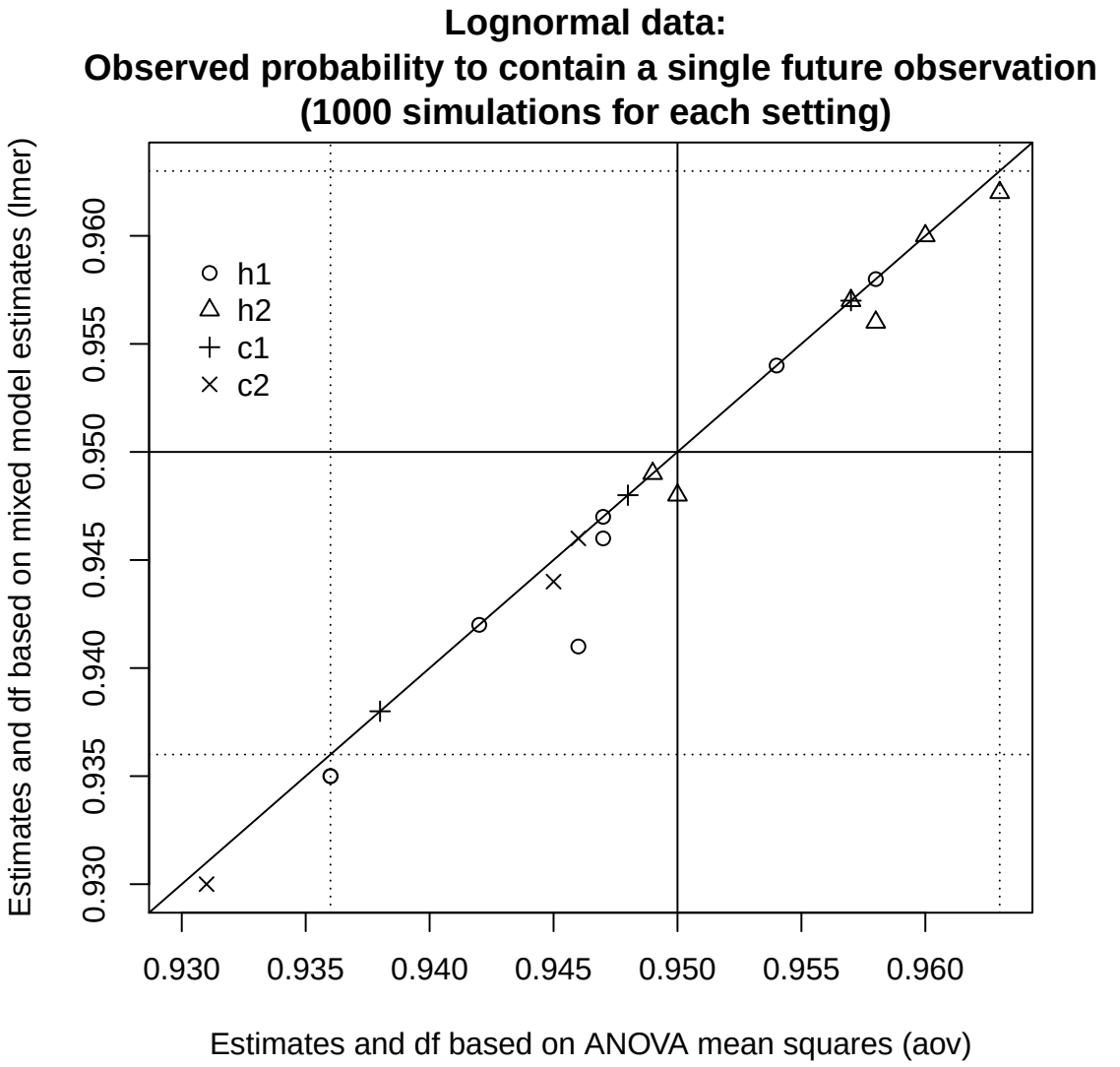

Figure 2: Observed proportion of cases where a single future observation is covered by a upper 0.95 prediction limit, for **log normal distributed data** from models h1, h2, c1, c2. Proportions estimated from 1000 simulation runs for each model and parameters setting, dotted lines circumscribe the range of observed proportions for which $H_0 : proportion \neq 0.95$ can not be rejected in a two-tailed test at level $\alpha 0.05$, i.e. if a method would be exact, 95% of the simulation results fall between the two dotted lines.

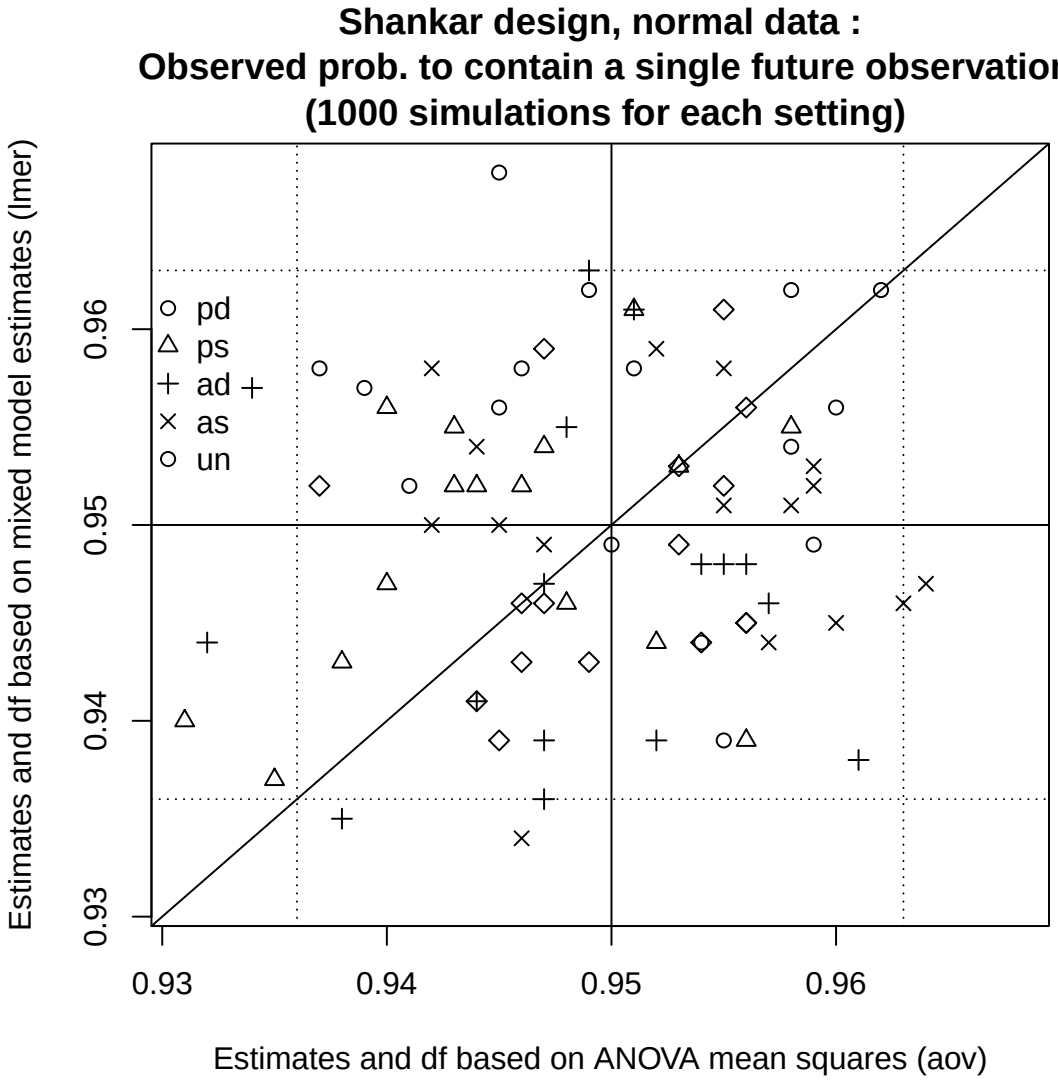

Figure 3: Observed proportion of cases where a single future observation is covered by a upper 0.95 prediction limit, for normal distributed data from **model ch3, i.e., the 'Shankar' design**. Proportions estimated from 1000 simulation runs for each parameters setting, dotted lines circumscribe the range of observed proportions for which $H_0 : proportion \neq 0.95$ can not be rejected in a two-tailed test at level $\alpha 0.05$, i.e. if a method would be exact, 95% of the simulation results fall between the two dotted lines. Strings **pd, ps, ad, as, un** correspond to different variance components settings: pd: patient dominating; ad: analyst + plate dominating; as: analyst+plate nearly absent, ps: patient nearly absent; un: nearly absent, unreliable technical device and analyst + noise, i.e., interaction variance components large

5 Open problems in application to ADA cutpoint definition

Prediction intervals for an empirically defined subpopulation

To include the uncertainty of classification, T. Jaki suggested to sample from the population of non-responders, depending the posterior probabilities to be a non-responder from the classification process. As an heuristic approach to include this uncertainty in the estimation of prediction intervals, one might include the posterior probabilities as weights during fitting the model. As a consequence, individuals from the overlap between responders and non-responders would be included when estimating the model for non-responders, but would be down-weighted depending on the estimated uncertainty of belonging to the non-responders. The background of the problem is the following:

The data sets available so far suggest, that the control population consists of two subpopulations (on the level of patient/subject/sample): a larger part of non-responders, with low mean, low variance and limited right-skewness, and a smaller subpopulation of responders, showing higher mean, higher variance and considerably right skewed distribution. Depending on the objective of applying the prediction interval, one might be interested in

- an upper $(1-\alpha)$ bound for the population of non-responders. This might be used in a screen for substances: if a single new substance exceeds this upper bound, it is considered hazardous. The probability to erroneously considering a non-hazardous substance to be hazardous is controlled via (α) of the upper bound for non-responders.

- a lower $(1-\alpha)$ bound for the population of responders. This might be used to classify a single new subjects: a single new subject might be considered a non-responder if it is below this lower bound. The probability to erroneously classify a responder as a non-responder, is controlled via (α) of the lower bound for the responders.

Applying the prediction intervals to define upper or lower bounds in this situations requires to distinguish the subpopulations of responders and non-responders. If the two populations show a considerable overlap (as was the case in the data sets analyzed so far), there is considerable uncertainty in assigning the upper tail of responders and the lower tail of non-responders. If such two subpopulations are divided by a simple cut-off, the upper tail of responders and lower tail of non-responders are too short. As a consequence, estimated variance components may be too small, and means of the non-responders and responders may be biased negatively and positively, respectively. Then, the prediction intervals will be liberal for those tails that are of interest in the decision.

Effect of responder/non-responder selection on balancedness of designs

Assume that patients are randomized w.r.t. further structures in the experimental design. A selection of responders and non-responders at the patient level and subsequent consideration of only the non-responders may cause hierarchical designs (h1, h2 or higher hierarchical) to be at least slightly unbalanced. Then, the above formulas for the estimation of prediction intervals are not directly applicable. If responder/non-responder selection takes place strictly at the patient level, and patients are crossed with the remaining effects, as in designs c1, c2, c2h1(Shankar), considering the non-responders only does not lead to unbalanced designs.

Inconsistency of software solutions for non-responders/responder selection and prediction intervals

For fitting mixture distributions so far only methods are available, that allow random effects at the patient level, and nesting structures within patients. With the crossed designs c1, c2, c2h1(Shankar), this assumption is violated. Using the available flexmix code might be justified after normalization on the levels of those random effects crossed with patients.

Prediction intervals for a mean of future technical replications

Consider models h2, or c2 which contain technical replications, e.g. $k = 2$ observations of, say, patient i at plate j. In a future process, a given patient $i*$ on plate $j*$ might again be analyzed with $K = 2$ technical replications. If the mean of the K future technical replications, $\bar{y}^* = \sum_k y_{i*j*k}/K$, is to be used for decision, a prediction interval for a future mean of the technical replicates would be appropriate.

For model $h2$, e.g., the variance term of such an interval, $\hat{V}(\hat{\mu}) + \hat{V}(\bar{y}^*)$ should than consist of the variance of the estimated general mean as above,

$$\hat{V}(\hat{\mu}) = (\hat{\sigma}_a^2/I, \hat{\sigma}_b^2/(IJ), \hat{\sigma}_e^2/(IJK)),$$

and the variance of a future mean of the K technical replicates,

$$\hat{V}(\bar{y}^*) = (\hat{\sigma}_a^2, \hat{\sigma}_b^2, \hat{\sigma}_e^2/K),$$

That is, while the patient and plate variance, $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$ contribute as above, the weight of the residual variance, σ_e^2, will decrease depending on K.

If this is correct, the above formulas can be used to estimate prediction intervals for future means of technical replicates: only the weights w_c for computing error terms and degrees of freedom have to be changed accordingly. Not yet implemented or validated.

6 Annex: Parameter settings of the simulation study

References

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.

Sahai, M. I. and Ageel, H. (2000). *The Analysis of Variance: Fixed, Random and Mixed Models*. Birkhäuser, Boston.

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