Quntification of difference in non-selectivity between compared IVD-MDs

Supplemental file

Pernille Kjeilen Fauskanger Sverre Sandberg Jesper Johansen Thomas Keller
Jeff Budd Greg Miller Vincent Delatour Bård Støve Anne Stavelin

August 26, 2022

Theory

Derivation of average prediction error variance for ordinary least squares regression

Derivation of average prediction error variance for Deming regression

The definition of $\zeta^\star$, depends on $\text{Var}[y_0 - \hat{y}_0]$, which we again must estimate, but instead of employing ordinary least squares regression to do so, we apply Deming regression. Using Deming regression based on (Fuller, Gillard), $\text{Var}[y_0 - \hat{y}_0]$ may theoretically be calculated by

$$\text{Var}[y_0 - \hat{y}_0] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{R_i}\sum_{r=1}^{R_i}\left[\text{Var}[b_1](x_{ir} - \overline{\overline{x}})^2 + \text{Var}[b_1]\sigma_{\text{MS}_X} + (1 + R_i^{-1}n^{-1})(\sigma_{\text{MS}_Y} + \beta_1\sigma_{\text{MS}_X})\right] \quad (1)$$

Then, assuming that $R_i = R \,\forall\, i \leq n$, this would collapse to

$$\begin{aligned}\text{Var}[y_0 - \hat{y}_0] &= \frac{\text{Var}[b_1]\sum_{i=1}^{n}\sum_{r=1}^{R}(x_{ir} - \overline{\overline{x}})^2}{nR} + \text{Var}[b_1]\sigma_{\text{MS}_X} + \frac{\beta_1^2\sigma_{\text{MS}_X} + \sigma_{\text{MS}_Y}}{nR} + \beta_1^2\sigma_{\text{MS}_X} + \sigma_{\text{MS}_Y} \\ &= \text{Var}[b_1]s_{XX} + \text{Var}[b_1]\sigma_{\text{MS}_X} + (1 + n^{-1}R^{-1})(\beta_1^2 + \Lambda)\sigma_{\text{MS}_X}, \end{aligned} \quad (2)$$

where an estimator for this is

$$\widehat{\text{Var}}[y_0 - \hat{y}_0] = \widehat{\text{Var}}[b_1]s_{XX} + \widehat{\text{Var}}[b_1]\widehat{\sigma_{\text{MS}_X}} + (1 + n^{-1}R^{-1})(\beta_1^2 + \lambda)\sigma_{\text{MS}_X} \quad (3)$$

Simulations

Simulation settings

The aim of the simulations is to confirm theoretical results, and possible reveal other fruitful relationships. Based on whether we want to hold the simulation parameters constant, or let them be randomly sampled, a subset of the steps one through five may be discarded in each simulation setting. Suppose we require 25 CSs measured in triplicate – this would imply having $n = 25$ and $R = 3$, signifying the two first points will be discarded. Alternatively, by setting the MS CVs to specific values (e.g., 2% and 1% for MS_X and MS_Y, respectively), the third point will be skipped. Or, we may e.g., fix the concentration interval to be between 100 and 200 units, meaning that fourth and fifth point get skipped. Note that these five first points are inspired by what we have seen for real data. In case the MS CVs, CV_{MS_X} for MS_X and CV_{MS_Y} for MS_X are randomly chosen, They are *independently* drawn from beta distributions with shape parameters 2 and 5, respectively. Because the sample space of $\text{Beta}(2, 5)$ is between $[0,\ 1]$, with mean $2/7 \approx 0.2857$, we divide

our observation from Beta$(2, 5)$ by 10, so that the sample space is mapped to $[0,\ 0.10]$, where realistic MS CVs are likely to be within. The lower and upper limits of the given concentration interval are defined by U_1 and $U_2 = U_1(1 + T)$, respectively, where U_1 and T are drawn continuous distributions matching reference intervals for numerous of analytes and satisfying that $U_1 \leq U_2$ theoretically.

Simulation setting 1

The following setting simulates values of ζ with concentration-independent CVs and MSs with identical selectivity profiles:

1. Draw the number of CSs, n, from a truncated Poisson distribution with cutoffs at 20 and 30, that is
$$n \sim \max\Big[20, \min\big[\text{Poisson}(25), 30\big]\Big] \tag{4}$$
2. Draw R from $\{2, 3, 4\}$ with respective point mass probabilities $(2/20, 17/20, 1/20)$.
3. Draw $\text{CV}_{\text{MS}_X}, \text{CV}_{\text{MS}_Y} \sim \text{Beta}(2, 5)/10$.
4. Draw $U_1 \sim F_{1.06,8.15} \cdot 44$ (F-distribution with 1.06 and 8.15 degrees of freedom, where the observation is scaled by 44). Then, draw $T \sim \text{Beta}(0.78, 11) \cdot 44$, and then calculate $U_2 = U_1(1 + T)$.
5. Set the lower range of the concentration range to U_1 and the upper range of the concentration range to U_2.
6. Simulate latent analyte concentrations, that we denote τ (tau) from a continuous uniform distribution defined for values within $[U_1, U_2]$, for all n CSs.
7. Calculate MS SDs (standard deviations) by using CVs simulated from either the third point or specific MS CVs, in addition to a randomly chosen concentration interval (i.e., following fourth and fifth point), or a particular concentration interval:
$$\sigma_{\text{MS}_X} = \text{CV}_{\text{MS}_X} \cdot \frac{1}{2}[U_1 + U_2] \tag{5}$$
$$\sigma_{\text{MS}_Y} = \text{CV}_{\text{MS}_Y} \cdot \frac{1}{2}[U_1 + U_2] \tag{6}$$
8. Add measurement error to latent analyte concentration (τ_i) for all $i = 1, 2, \ldots, n-1, n$:
$$x_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\text{MS}_X}), \forall r = 1, \ldots R \tag{7}$$
$$y_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\text{MS}_Y}), \forall r = 1, \ldots R \tag{8}$$
The simulated values x_{ir} represents measurements based on MS_X, and $\{y_{ir}\}$ signify measurements from MS_Y.
9. Calculate SD_{MS_X}, SD_{MS_Y}, and then λ, using the simulated observed measurements from 8.
10. Calculate $S^2_{\text{P}_{\text{AR}}}$ and the b_1, that is the regression slope coefficient, using simulated observed measurements from point eight and λ from the ninth point.
11. Calculate ζ

Simulation setting 2

We define η to be the so-called *heteroscedasticity factor*. In addition, we denote η_0 to be the *proportion of base* MS standard deviations (base MS standard deviations are e.g., σ_{MS_X}, σ_{MS_Y}, or $\sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}}$). Having $\eta \geq \eta_0$, entails that MS SDs will gradually increase from the lower part of the concentration interval to the upper. In contrast, having $\eta_0 \geq \eta$ implies that MS SDs will decrease from the lower part of the concentration interval to the upper. Suppose that $\eta = 2 \cdot \eta_0$. In this case, we would expect that MS SDs increase from $\text{base} \cdot \eta_0$ to $2 \cdot \text{base} \cdot \eta_0$ over the concentration interval. Conversely, having $\eta = 1/2 \cdot \eta_0$, we expect the MS SDs to decline from $\text{base} \cdot \eta_0$ to $\frac{1}{2} \cdot \text{base} \cdot \eta_0$ across the concentration interval. So, now that the heteroscedasticity factor and proportion of base MS SDs are defined, we can start describing the second simulation setting for ζ.

1. Draw the number of CSs, n, from a truncated Poisson distribution with cutoffs at 20 and 30, that is
$$n \sim \max\Big[20, \min\big[\text{Poisson}(25), 30\big]\Big] \tag{9}$$

2. Draw R from $\{2,\ 3,\ 4\}$ with respective point mass probabilities $(^2/_{20}, ^{17}/_{20}, ^1/_{20})$.
3. Draw $\text{CV}_{\text{MS}_X}, \text{CV}_{\text{MS}_Y} \sim \text{Beta}(2,5)/10$.
4. Draw $U_1 \sim F_{1.06,8.15} \cdot 44$ (F-distribution with 1.06 and 8.15 degrees of freedom, where the sample is scaled by 44). Then, draw $T \sim \text{Beta}(0.78, 11) \cdot 44$, and then calculate $U_2 = U_1(1+T)$.
5. Set the lower range of the concentration range to U_1 and the upper range of the concentration range to U_2.
6. Simulate latent analyte concentrations, that we denote τ (tau) from a continuous uniform distribution defined for values within $[U_1, U_2]$, for all n CSs. Thereafter, sort $\tau_1, \tau_2, \ldots, \tau_n$ in ascending order. The sorted values will be denoted $\tau_{(1)}, \tau_{(2)}, \ldots, \tau_{(n)}$.
7. Calculate MS SDs by using CVs simulated from either the third point or specific MS CVs, in addition to a randomly chosen concentration interval (i.e., following fourth and fifth point), or a particular concentration interval:

$$\sigma_{\text{MS}_X} = \text{CV}_{\text{MS}_X} \cdot \frac{1}{2}[U_1 + U_2] \tag{10}$$

$$\sigma_{\text{MS}_Y} = \text{CV}_{\text{MS}_Y} \cdot \frac{1}{2}[U_1 + U_2] \tag{11}$$

8. Set $\text{base} = \sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}}$. Then, select n equally spaced points within the interval,

$$[\eta_0 \cdot \text{base},\ \eta \cdot \eta_0 \cdot \text{base}], \tag{12}$$

where the first and last point being $\eta_0 \cdot \text{base}$ and $\eta \cdot \eta_0 \cdot \text{base}$, respectively. We will denote the this set of points by $\{P_i\}$, defined by,

$$\{P_i\} = \Big\{\eta_0 \cdot \text{base}, \ldots, \eta \cdot \eta_0 \cdot \text{base}\Big\} \tag{13}$$

9. Add measurement error to ascending latent analyte concentration $(\tau_{(i)})$ for all $i = 1, \ldots, n$:

$$x_{ir} = \tau_{(i)} + \mathcal{N}(0, P_i^2),\ \forall\ r = 1, \ldots R \tag{14}$$

$$y_{ir} = \tau_{(i)} + \mathcal{N}(0, P_i^2),\ \forall\ r = 1, \ldots R \tag{15}$$

The simulated values $\{x_{ir}\}$ represents measurements based on MS_X, and $\{y_{ir}\}$ signify measurements from MS_Y.

10. Calculate SD_{MS_X}, SD_{MS_Y}, and then λ, using the simulated observed measurements from point eight.
11. Calculate $S^2_{\text{P}_{\text{AR}}}$ and the b_1, that is the regression slope coefficient, using simulated observed measurements from point nine and λ from the 10. point.
12. Calculate ζ

Simulation setting 3

Let p denote the theoretical probability of a CS getting affected by having *random dins* between compared MSs. In this context, affected by dins means that the cluster of measurements for the CS is relocated in the XY-plane directed away from the true regression line suggesting the relationship between MS_X and MS_Y, by a random magnitude between 0 and a *particular maximum relocation magnitude*, referred to by $m_{\max}$. Setting $p > 0$, $m_{\max} > 0$ will namely introduce dins in simulated data, and the grade of dins is proportional to both p and $m_{\max}$. In order to generalize the principle of maximum relocation magnitude, we must talk about relative maximum relocation magnitude, which we define as a MS SD base (e.g., σ_{MS_X}, σ_{MS_Y} or

$\sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}}$) scaled by a real number. Suppose we use $\sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}}$ to be the MS SD base. Then, if we choose a value $m_{\max} \in \mathbb{R}$, the maximum relocation magnitude would be

$$m_{\max} \cdot \text{base} = m_{\max} \cdot \sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}}. \tag{16}$$

If CS i is affected by dins with original non-affected measurements, $y_{i1}, \ldots, y_{iR}$, the sample space for the affected measurements, $y^\star_{i1}, \ldots, y^\star_{iR}$ is

$$\Omega_{y^\star_{ir}} = \Big[y_{ir},\ y_{ir} + m_{\max} \cdot \text{base}\Big] = y_{ir} + \Big[0,\ m_{\max} \cdot \text{base}\Big],\ \forall\ r = 1, \ldots, R \tag{17}$$

The random variable Y_i that is the relocation magnitude supported between 0 and $m_{\max}$·base, may be simulated by a distribution that supports values on $[0, m_{\max} \cdot \text{base}]$ or a distribution with support $[0, \infty)$ truncated at $m_{\max} \cdot \text{base}$. The Beta-distribution is a natural choice to simulate from. Indeed, $\text{Beta}(2,2) \cdot m_{\max} \cdot \text{base}$ is supported on $[0, m_{\max} \cdot \text{base}]$, and is symmetric around $1/2 \cdot m_{\max} \cdot \text{base}$ making it favorable. Simulation setting 3 is defined by the following algorithm:

1. Draw the number of CSs, n, from a truncated Poisson distribution with cutoffs at 20 and 30, that is

$$n \sim \max\Big[20, \min\Big[\text{Poisson}(25), 30\Big]\Big] \tag{18}$$

2. Draw R from $\{2, 3, 4\}$ with respective point mass probabilities $(2/20, 17/20, 1/20)$.
3. Draw $\text{CV}_{\text{MS}_X}, \text{CV}_{\text{MS}_Y} \sim \text{Beta}(2, 5)/10$.
4. Draw $U_1 \sim F_{1.06, 8.15} \cdot 44$ (F-distribution with 1.06 and 8.15 degrees of freedom, where the observation is scaled by 44). Then, draw $T \sim \text{Beta}(0.78, 11) \cdot 44$, and then calculate $U_2 = U_1(1 + T)$.
5. Set the lower range of the concentration range to U_1 and the upper range of the concentration range to U_2.
6. Simulate latent analyte concentrations, that we denote τ (tau) from a continuous uniform distribution defined for values within $[U_1, U_2]$, for all n CSs.
7. Calculate MS SDs (standard deviations) by using CVs simulated from either the third point or specific MS CVs, in addition to a randomly chosen concentration interval (i.e., following fourth and fifth point), or a particular concentration interval:

$$\sigma_{\text{MS}_X} = \text{CV}_{\text{MS}_X} \cdot \frac{1}{2}[U_1 + U_2] \tag{19}$$

$$\sigma_{\text{MS}_Y} = \text{CV}_{\text{MS}_Y} \cdot \frac{1}{2}[U_1 + U_2] \tag{20}$$

8. Add measurement error to latent analyte concentration (τ_i) for all $i = 1, 2, \ldots, n-1, n$:

$$x_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\text{MS}_X}), \forall r = 1, \ldots R \tag{21}$$

$$y_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\text{MS}_Y}), \forall r = 1, \ldots R \tag{22}$$

   The simulated values $\{x_{ir}\}$ represents original measurements based on MS_X, and $\{y_{ir}\}$ signify original measurements from MS_Y.

9. Draw the subscripts referring to the affected CSs that are affected by dins between compared MSs:
   1. Draw $X \sim \text{binomial}(n, p)$.
   2. Draw X observations uniformly from $\{1, 2, \ldots, n\}$. The random sample is denoted by $\{X_j\}$. It is possible that $\{X_j\} = \emptyset$, which happens when observed X is zero.
   3. Draw X relocation magnitude observations, that we will call $\{Y_j\}$, from $\text{Beta}(2,2) \cdot m_{\max} \cdot \text{base}$ (skip if $\{X_j\} = \emptyset$).

4. Draw X direction observations, that we will call $\{D_j\}$, from $1 - 2 \cdot \text{binomial}(1,\ 1/2)$ (skip if $\{X_j\} = \emptyset$).

Add random difference in non-selectivity effects to the original measurement results of MS_Y if $\{X_j\} \neq \emptyset$ and $i \in \{X_j\}$ (otherwise, skip):

$$y^{\star}_{X_j,\,r} = y_{X_j,\,r} + Y_j \cdot D_j,\ \forall\, r = 1, \ldots R \tag{23}$$

10. Calculate SD_{MS_X}, SD_{MS_Y}, and then λ, using the simulated affected measurements from the ninth point.
11. Calculate $S^2_{\text{P}_{\text{AR}}}$ and the b_1, that is the regression slope coefficient, using simulated observed measurements from point eight and λ from the ninth point.
12. Calculate ζ

Simulation setting 4

Let $q_{l,u}$ denote the *quantile range* (with width $u - l$), that is a subset of $[0,\ 1]$. The quantile range correspond with a subset of the true analyte concentrations, $\{\tau_i\}$. We denote the τ values corresponding with a given $q_{l,u}$ by $\{Q_j\}$. Mathematically, the CSs affected by *systematic dins* are defined by $\{Q_j\} = \{\tau_i : l \leq \text{P}(\tau_i \leq \tau) \leq u\}$. For example, if $\{l = 0, u = 0.25\}$, the subset of $\{\tau_i\}$ values satisfying $\text{P}(\tau_i \leq \tau) \leq 0.25$ will be selected. Alternatively, if $\{l = 0.70, u = 1.00\}$, particular values of $\{\tau_i\}$ solving $0.70 \leq \text{P}(\tau_i \leq \tau) \leq 1.00$ are selected. The τ values selected from $\{\tau_i\}$, $\{Q_j\}$, will be relocated in the XY-plane away from the regression line defining the relationship between measurements of MS_X and MS_Y in a similar, but not identical way, to the relocation rules of simulation setting 3. In contrast to simulation setting 3, the actual relocation magnitudes are not random, but deterministic. The relocation magnitudes will not be fixed, but rather systematic increasing over towards the boundaries if $\{l = 0, u > 0\}$ or $\{l < 1, u = 1\}$, which are the most realistic situations. $\{l = 0, u > 0\}$ suggests systematic dins in the lower range of the concentration interval, and $\{l < 1, u = 1\}$ indicates systematic dins in the upper range of the concentration interval. As with simulation setting 3, we will also here consider a MS SD base, that $m_{\max}$ is scaled with, making the relocation magnitudes relative. Q_j is relocated with the following magnitude:

$$Y_j = m_{\max} \cdot \sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}} \begin{cases} \dfrac{U_1 + (u-l)\cdot(U_2 - U_1) - Q_{(j)}}{U_1 + (u-l)\cdot(U_2 - U_1)}, & \text{for } l = 0, u > 0 \\ \dfrac{Q_{(j)} - U_2 + (u-l)\cdot(U_2 - U_1)}{(u-l)\cdot(U_2 - U_1)}, & \text{for } l < 1, u = 1 \end{cases}$$

The relocation direction may be either above or below the regression line, and the direction (1 for above and -1 for below) is simulated from $1 - 2 \cdot \text{binomial}(1, 1/2)$. **Figure 4X** illustrates how this relocation framework operates for $\{l = 0, u = 0.25\}$ and $\{l = 0.70, u = 1\}$ having $U_1 = 50, U_2 = 100, \text{CV}_{\text{MS}_X} = 2\%, \text{CV}_{\text{MS}_Y} = 3\%, n = 50$, and $R = 3$. With the principles covered, we simulate ζ values base on simulation setting 4 by following the algorithm below:

1. Draw the number of CSs, n, from a truncated Poisson distribution with cutoffs at 20 and 30, that is

$$n \sim \max\Big[20, \min\big[\text{Poisson}(25), 30\big]\Big] \tag{24}$$

2. Draw R from $\{2, 3, 4\}$ with respective point mass probabilities $(2/20, 17/20, 1/20)$.
3. Draw $\text{CV}_{\text{MS}_X}, \text{CV}_{\text{MS}_Y} \sim \text{Beta}(2, 5)/10$.
4. Draw $U_1 \sim F_{1.06,8.15} \cdot 44$ (F-distribution with 1.06 and 8.15 degrees of freedom, where the observation is scaled by 44). Then, draw $T \sim \text{Beta}(0.78, 11) \cdot 44$, and then calculate $U_2 = U_1(1 + T)$.
5. Set the lower range of the concentration range to U_1 and the upper range of the concentration range to U_2.

6. Simulate latent analyte concentrations, that we denote τ (tau) from a continuous uniform distribution defined for values within $[U_1, U_2]$, for all n CSs.

7. Calculate MS SDs (standard deviations) by using CVs simulated from either the third point or specific MS CVs, in addition to a randomly chosen concentration interval (i.e., following fourth and fifth point), or a particular concentration interval:

$$\sigma_{\mathrm{MS}_X} = \mathrm{CV}_{\mathrm{MS}_X} \cdot \frac{1}{2}[U_1 + U_2] \tag{25}$$

$$\sigma_{\mathrm{MS}_Y} = \mathrm{CV}_{\mathrm{MS}_Y} \cdot \frac{1}{2}[U_1 + U_2] \tag{26}$$

8. Add measurement error to latent analyte concentration (τ_i) for all $i = 1, 2, \ldots, n-1, n$:

$$x_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\mathrm{MS}_X}),\ \forall\ r = 1, \ldots R \tag{27}$$

$$y_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\mathrm{MS}_Y}),\ \forall\ r = 1, \ldots R \tag{28}$$

   The simulated values $\{x_{ir}\}$ represents original measurements based on MS_X, and $\{y_{ir}\}$ signify original measurements from MS_Y.

9. Obtain the subscripts referring to the CSs affected by systematic dins between compared MSs:

   1. Draw $D \sim 1 - 2 \cdot \mathrm{binomial}(1,\ 1/2)$
   2. Obtain $\{X_j\} = \{i :\ l \leq \mathrm{P}(\tau_i \leq \tau) \leq u\}$
   3. Calculate Y_j for $j = 1, \ldots, |\{X_j\}|$. Here, $|\{X_j\}|$ is the cardinality of $\{X_j\}$.

   Add the systematic difference in non-selectivity effects to the original measurement results of MS_Y if $\{X_j\} \neq \emptyset$ and $i \in \{X_j\}$ (otherwise, skip):

$$y^{\star}_{X_j, r} = y_{X_j, r} + Y_j \cdot D,\ \forall\ r = 1, \ldots R \tag{29}$$

10. Calculate $\mathrm{SD}_{\mathrm{MS}_X}$, $\mathrm{SD}_{\mathrm{MS}_Y}$, and then λ, using the simulated affected measurements from the ninth point.

11. Calculate $S^2_{\mathrm{P_{AR}}}$ and the b_1, that is the regression slope coefficient, using simulated observed measurements from point eight and λ from the ninth point.

12. Calculate ζ

Simulation setting 5

Suppose we have two MS comparisons that are nearly identical. However, in the first comparison, the MSs do not have dins, but the second comparison do. Other than their difference on dins, the two MS comparisons are identical. We estimate point-wise prediction intervals for $\{y_{ir}\}$ based on $\{x_{ir}\}$ for both MS comparisons using an ordinary least squares linear model, yielding two sets of prediction interval widths. Taking the average of the squared elements within each set of prediction interval widths, generates two quantities, that we name w_0 and w. w_0 is the average squared width of the point-wise prediction intervals for the MS comparison without dins, which entails that w is the average squared width of the point-wise prediction intervals for the MS comparison with non-zero dins. Based on the ordinary least squares linear model the expressions for w_0 and w are

$$w_0 = \frac{4t^2 \cdot S^2_1}{nR}(nR + 2) \tag{30}$$

$$w = \frac{4t^2 \cdot S^2_2}{nR}(nR + 2) \tag{31}$$

Here, t is a quantile of the $t(nR - 2)$ distribution defined by the confidence level for the point-wise prediction intervals. The exact value of t is not important. We now define the square-root of the ratio of w and w_0 by

$1 + M$:

$$1 + M = \sqrt{\frac{w}{w_0}} = \sqrt{\frac{S_2^2(nR+2)nR}{S_1^2(nR+2)nR}} \tag{32}$$

We can divide by $\text{SD}^2_{\text{MS}_Y} + b_1^2\text{SD}^2_{\text{MS}_X}$ or $b_1^2\text{SD}^2_{\text{MS}_Y} + \text{SD}^2_{\text{MS}_X}$ in both numerator and denominator which gives results in

$$1 + M = \sqrt{\frac{\zeta}{\zeta_0}} \Rightarrow \sqrt{\zeta} = (1 + M)\sqrt{\zeta_0} \tag{33}$$

This equation implicitly tells us that M is the relative percentage increase of the root of the average squared point-wise prediction interval widths between the first and second MS comparison. Assuming that only a subset of $\{y_{ir}\}$ (i.e., none of $\{x_{ir}\}$) are relocated stemming from dins, M may also be interpreted as the relative percentage increase of the average point-wise prediction interval widths between the first and second MS comparison. Squaring both sides of the equation yields a closed-form relationship between a ζ value calculated based on any degree of dins, and a ζ value calculated based on a MS comparison without dins (ζ_0):

$$\zeta = (1 + M)^2\zeta_0 \tag{34}$$

We can then simulate ζ values affected by dins between compared MSs by choosing $M > 0$. The algorithm defining the fifth simulation setting is:

1. Draw the number of CSs, n, from a truncated Poisson distribution with cutoffs at 20 and 30, that is

$$n \sim \max\left[20, \min\left[\text{Poisson}(25), 30\right]\right] \tag{35}$$

2. Draw R from $\{2, 3, 4\}$ with respective point mass probabilities $(2/20, 17/20, 1/20)$.
3. Draw $\text{CV}_{\text{MS}_X}, \text{CV}_{\text{MS}_Y} \sim \text{Beta}(2, 5)/10$.
4. Draw $U_1 \sim F_{1.06,8.15} \cdot 44$ (F-distribution with 1.06 and 8.15 degrees of freedom, where the observation is scaled by 44). Then, draw $T \sim \text{Beta}(0.78, 11) \cdot 44$, and then calculate $U_2 = U_1(1 + T)$.
5. Set the lower range of the concentration range to U_1 and the upper range of the concentration range to U_2.
6. Simulate latent analyte concentrations, that we denote τ (tau) from a continuous uniform distribution defined for values within $[U_1, U_2]$, for all n CSs.
7. Calculate MS SDs (standard deviations) by using CVs simulated from either the third point or specific MS CVs, in addition to a randomly chosen concentration interval (i.e., following fourth and fifth point), or a particular concentration interval:

$$\sigma_{\text{MS}_X} = \text{CV}_{\text{MS}_X} \cdot \frac{1}{2}[U_1 + U_2] \tag{36}$$

$$\sigma_{\text{MS}_Y} = \text{CV}_{\text{MS}_Y} \cdot \frac{1}{2}[U_1 + U_2] \tag{37}$$

8. Add measurement error to latent analyte concentration (τ_i) for all $i = 1, 2, \ldots, n-1, n$:

$$x_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\text{MS}_X}), \forall r = 1, \ldots R \tag{38}$$

$$y_{ir} = \tau_i + \mathcal{N}(0, \sigma^2_{\text{MS}_Y}), \forall r = 1, \ldots R \tag{39}$$

   The simulated values x_{ir} represents measurements based on MS_X, and $\{y_{ir}\}$ signify measurements from MS_Y.
9. Calculate SD_{MS_X}, SD_{MS_Y}, and then λ, using the simulated observed measurements from 8.
10. Calculate $S^2_{\text{P}_{\text{AR}}}$ and the b_1, that is the regression slope coefficient, using simulated observed measurements from point eight and λ from the ninth point.
11. Calculate ζ_0 and use it together with M to calculate ζ.

Simulation setting 6

Much the same as above, but heteroscedasticity is added in addition:

1. Draw the number of CSs, n, from a truncated Poisson distribution with cutoffs at 20 and 30, that is
$$n \sim \max\Big[20, \min\Big[\text{Poisson}(25), 30\Big]\Big] \tag{40}$$
2. Draw R from $\{2,\ 3,\ 4\}$ with respective point mass probabilities $(2/20, 17/20, 1/20)$.
3. Draw $\text{CV}_{\text{MS}_X}, \text{CV}_{\text{MS}_Y} \sim \text{Beta}(2,5)/10$.
4. Draw $U_1 \sim F_{1.06,8.15} \cdot 44$ (F-distribution with 1.06 and 8.15 degrees of freedom, where the sample is scaled by 44). Then, draw $T \sim \text{Beta}(0.78, 11) \cdot 44$, and then calculate $U_2 = U_1(1+T)$.
5. Set the lower range of the concentration range to U_1 and the upper range of the concentration range to U_2.
6. Simulate latent analyte concentrations, that we denote τ (tau) from a continuous uniform distribution defined for values within $[U_1, U_2]$, for all n CSs. Thereafter, sort $\tau_1, \tau_2, \ldots, \tau_n$ in ascending order. The sorted values will be denoted $\tau_{(1)}, \tau_{(2)}, \ldots, \tau_{(n)}$.
7. Calculate MS SDs by using CVs simulated from either the third point or specific MS CVs, in addition to a randomly chosen concentration interval (i.e., following fourth and fifth point), or a particular concentration interval:
$$\sigma_{\text{MS}_X} = \text{CV}_{\text{MS}_X} \cdot \frac{1}{2}[U_1 + U_2] \tag{41}$$
$$\sigma_{\text{MS}_Y} = \text{CV}_{\text{MS}_Y} \cdot \frac{1}{2}[U_1 + U_2] \tag{42}$$
8. Set $\text{base} = \sqrt{\sigma^2_{\text{MS}_X} + \sigma^2_{\text{MS}_Y}}$. Then, select n equally spaced points within the interval,
$$[\eta_0 \cdot \text{base},\ \eta \cdot \eta_0 \cdot \text{base}], \tag{43}$$
where the first and last point being $\eta_0 \cdot \text{base}$ and $\eta \cdot \eta_0 \cdot \text{base}$, respectively. We will denote the this set of points by $\{P_i\}$, defined by,
$$\{P_i\} = \Big\{\eta_0 \cdot \text{base}, \ldots, \eta \cdot \eta_0 \cdot \text{base}\Big\} \tag{44}$$
9. Add measurement error to ascending latent analyte concentration $(\tau_{(i)})$ for all $i = 1, \ldots, n$:
$$x_{ir} = \tau_{(i)} + \mathcal{N}(0, P_i^2),\ \forall\ r = 1, \ldots R \tag{45}$$
$$y_{ir} = \tau_{(i)} + \mathcal{N}(0, P_i^2),\ \forall\ r = 1, \ldots R \tag{46}$$
The simulated values $\{x_{ir}\}$ represents measurements based on MS_X, and $\{y_{ir}\}$ signify measurements from MS_Y.
10. Calculate SD_{MS_X}, SD_{MS_Y}, and then λ, using the simulated observed measurements from point eight.
11. Calculate $S^2_{\text{P}_{\text{AR}}}$ and the b_1, that is the regression slope coefficient, using simulated observed measurements from point nine and λ from the 10. point.
12. Calculate ζ_0 and use it together with M to calculate ζ.

Principle figures for the different simulation settings

Principle figures for simulation setting 2

Simulation parameters

We are interested in performing seven sets of simulations. The two first uses simulation setting 1, and the three, four, five, six and seven uses simulation settings two, three, four, five and six.

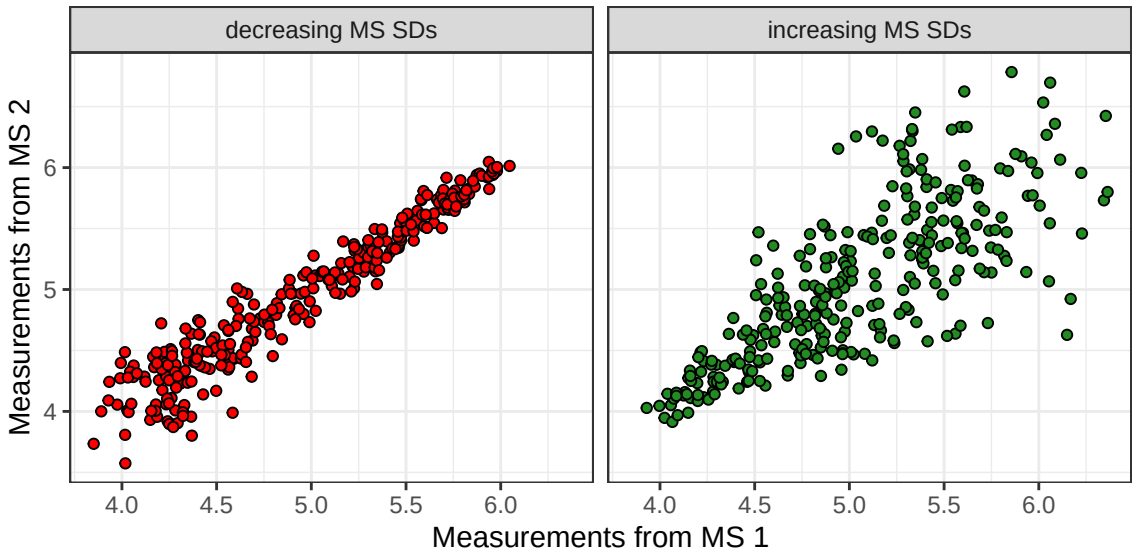


Figure 1: Visualization of heteroscedasticity with eta = 6 (top graph) and eta = 0.25 (bottom graph).

Simulation parameters in the first set of simulations

In the first set of simulations we are simulation ζ values with all combinations of

1. Concentration intervals:
   $\{U_1 = 5, U_2 = 10\}$, $\{U_1 = 2, U_2 = 10\}$, $\{U_1 = 500, U_2 = 750\}$ $\{U_1 = 70, U_2 = 1600\}$.
2. MS CVs for CV_{MS_X} and CV_{MS_Y}:
   0.1%, 0.5%, 1.0%, 2.5%, 5.0%, 7.5%, and 10%.
3. Number of clinical samples: 20 - 30
4. Number of replicated measurements: 2 - 4

100,000 ζ values are sampled for every combination of simulation parameters. There are 196 unique combinations of simulation parameters, which means that 19.6 million ζ values are simulated in total for these simulations.

Simulation parameters in the second set of simulations

In the second set of simulations we are simulation ζ values, based on simulation setting 1, with all combinations of

1. Number of clinical samples: 20, 25, 30, 35, 40
2. Number of replicated measurements: 2, 3, 4

MS CVs and concentration intervals are randomly sampled from the distributions given in simulation setting 1. 1,000,000 ζ values are sampled for every combination of simulation parameters. There are 15 unique combinations of simulation parameters, which means that 15 million ζ values are simulated for these simulations.

Simulation parameters in the third set of simulations

In the third set of simulations we are simulating ζ values, based on simulation setting 3, utilizing all combinations of

1. Number of clinical samples: 20, 25, 30, 35, 40
2. Number of replicated measurements: 2, 3, 4
3. Heteroscedasticity factors, η: 0.25, 0.50, 0.75, 1.5, 2, 4, 6

4. Proportion of base MS standard deviations: $\eta_0 = 1$

MS CVs and concentration intervals are randomly sampled from the distributions given in simulation setting 3. 250,000 ζ values are sampled for every combination of simulation parameters. There are 105 unique combinations of simulation parameters, which means that 26.25 million ζ values are simulated for these simulations.

Simulation parameters in the fourth set of simulations

In the fourth set of simulations we are simulating measurements from two MSs in a comparisons, that have random differences in non-selectivity. This implies that we wll simulate ζ values based on simulation setting 3. Accordingly the two parameters p (average proportion of CSs' measurements affected by random differences in non-selectivity) and $m_{\max}$ (maximal relocation magnitude of random dins-affected CSs' measurements) are included in addition to the usual 15 study designs. We simulate values of ζ using all combinations of

1. Number of clinical samples: 20, 25, 30, 35, 40
2. Number of replicated measurements: 2, 3, 4
3. Average proportion of affected CSs: 0.05, 0.10, 0.15, 0.20, 0.25, 0.30
4. Maximal relocation magnitudes: 1, 2, 3, 5, 7.5, 10

MS CVs and concentration intervals are randomly sampled from the distributions given in simulation setting 4. 250,000 ζ values are sampled for every combination of simulation parameters. There are 540 unique combinations of simulation parameters, which means that 0.125 billion ζ values are simulated in this simulation setting.

Simulation parameters in the fifth set of simulations

In the fifth set of simulations we are simulating measurements from two MSs in a comparisons, that have systematic differences in non-selectivity. This implies that we wll simulate ζ values based on simulation setting 4. Accordingly the two parameters q (quantile range where boundaries are either 0 or 1) and $m_{\max}$ (maximal relocation magnitude of systematic dins-affected CSs' measurements) are included in addition to the usual 15 study designs. We simulate values of ζ using all combinations of

1. Number of clinical samples: 20, 25, 30, 35, 40
2. Number of replicated measurements: 2, 3, 4
3. quantile ranges: (0, 0.05), (0, 0.10), (0, 0.15), (0, 0.20), (0, 0.25), (0.95, 1), (0.90, 1), (0.85, 1), (0.80, 1), (0.75, 1)
4. Maximal relocation magnitudes: 1, 2, 3, 5, 7.5, 10

MS CVs and concentration intervals are randomly sampled from the distributions given in simulation setting 4. 250,000 ζ values are sampled for every combination of simulation parameters. There are 1080 unique combinations of simulation parameters, which means that 0.250 billion ζ values are simulated in this simulation setting.

Simulation parameters in the sixth set of simulations

In the sixth set of simulations, we are simulating ζ values, based on simulation setting 5, which describes the more general approach to simulate ζ values corresponding with effects caused by differences in non-selectivity. Here we include the variable M, that is directly related to the increase of the average prediction interval width due to differences in non-selectivity. We will utilize the following parameter combinations:

1. Number of clinical samples: 20, 25, 30, 35, 40
2. Number of replicated measurements: 2, 3, 4
3. M values: $0.05, 0.06, \ldots, 0.99, 1.00$

MS CVs and concentration intervals are randomly sampled from the distributions given in simulation setting 5. 250,000 ζ values are sampled for every combinations of simulation parameters. There are 1440 unique combinations of simulation parameters, which means that 0.360 billion ζ values are simulated in this simulation

setting. We will use these simulation results to calculate the power of procedure we use to detect excessive differences in non-selectivity

Simulation results

Simulation results for the first set of simulations

A sequence of percentiles as well as the fourth first moments are calculated based on the raw simulation results. The first four moments are mean, variance, skewness and kurtosis. Skewness is interpreted as a measure of asymmetry for the distribution of ζ. In other words, how far is the distribution from a symmetric distribution. Negative skewness signify that the distribution is left-skewed, and positive skewness signify that it is right-skewed. The magnitude of the skewness estimate, quantifies the degree of the skew. Kurtosis is a measure of how thick the tails of the distribution are. Note that extreme ζ values are stripped here because the estimates of mean, variance, skewness and kurtosis are parametric. In this set of simulations, the actual values of ζ's moments is not important, as we just desire to prove the lack of relationship between ζ and MS CVs and concentration intervals. However, in the next set of simulations, there will be less stripping of extreme ζ values which signify that the moments estimates may differ considerably.

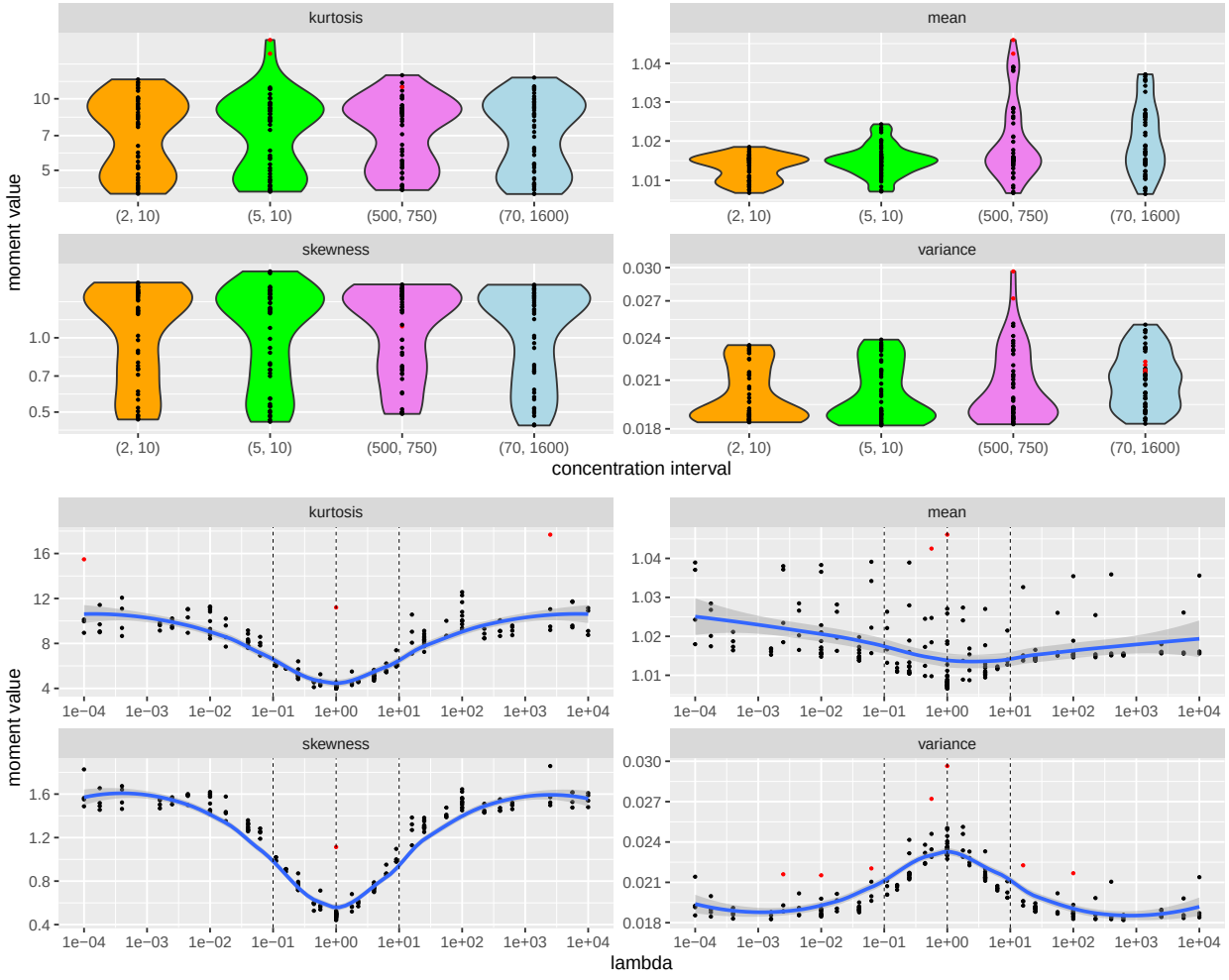


Figure 2: Empirical moments of zeta values versus concentration interval and lambda. The dashed lines signify lamba values 0.1, 1, and 10, respectively.

Based on figure 2, we conclude that the four first moments of ζ are practically independent of concentration intervals. This fact suggests that the distribution of ζ very likely have the approximately same form for

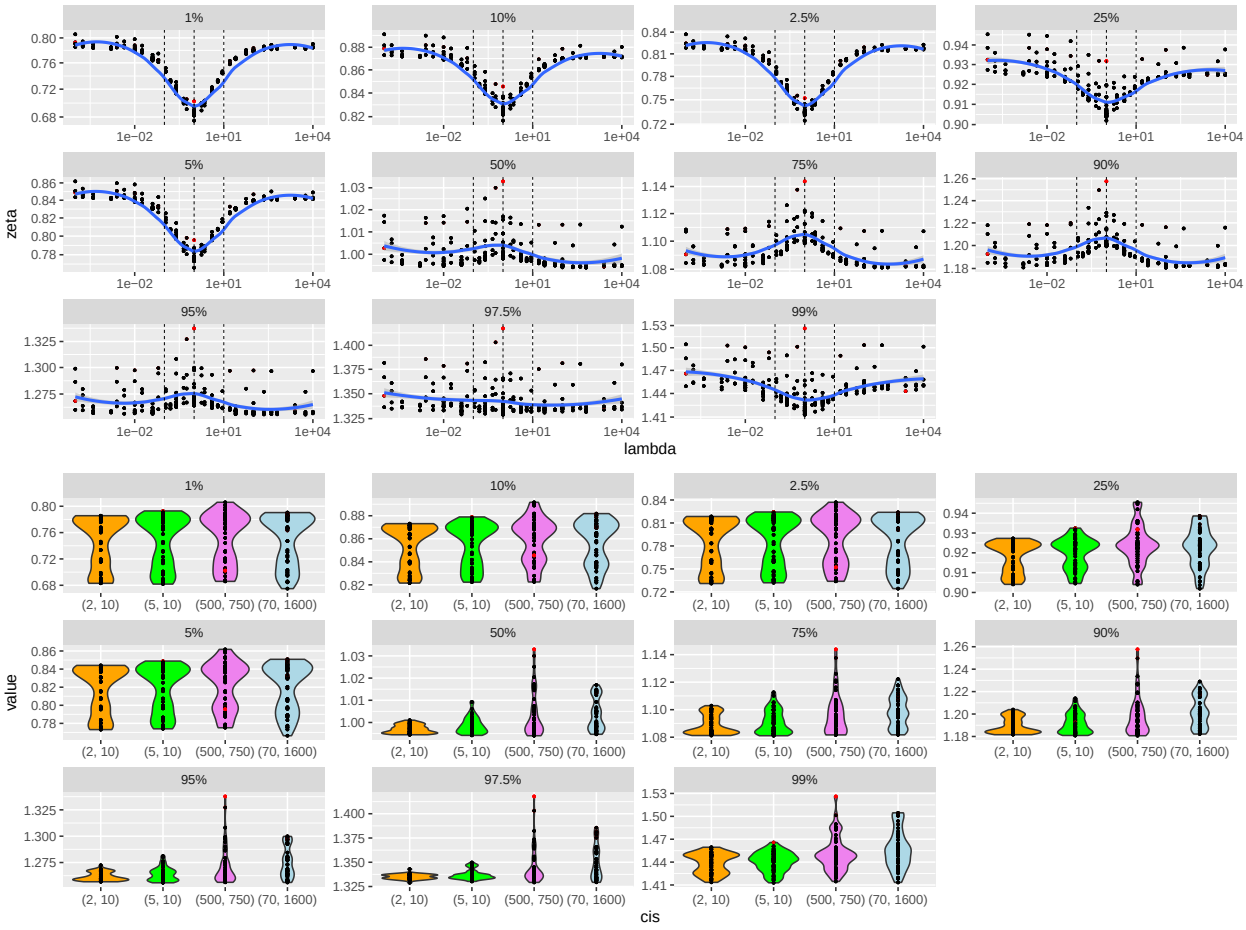


Figure 3: Empirical quantiles of zeta values versus concentration interval and lambda

most realistic concentration intervals. There are a couple of values (red points), that are somewhat extreme according to the IQR outlier test. These extremities are typically observed when λ is close to 1 for considerable values for both CV_{MS_X} and CV_{MS_Y} (e.g., CVs over 10.0%). From figure 3, we see that there is an interesting, but rather unimportant relationship between λ and each percentile of ζ. For the lower percentiles, that is, below 50%, most simulated percentiles seem to decrease as $\lambda \to 1$ from either side. For the upper percentiles, that is, above 50%, most simulated quantiles seem to increase as $\lambda \to 1$ from either side (except for the 99th percentile). However, the absolute difference between minimum and maximum for all these curves are negligible, which is why this relationship is unimportant. To summarize, the impact of the separate values of CV_{MS_X} and CV_{MS_Y} is minor. λ values close to 1 may potentially have an effect on ζ, but this effect is not general and small enough to ignore. To conclude, ζ is independent of concentration intervals, MS CVs and λ. From here, we will therefore not include particular values for MS CVs and concentration intervals, and leave those values to be randomly sampled. Thus, none of the points are skipped in simulation settings on a general basis.

Simulation results for the second set of simulations

In the second set of simulations, we will consider particular study designs and their impact on the distribution of ζ. We expect that small study designs yield more uncertainty in ζ compared to larger study designs. However, we do not expect that ζ is likely to be larger than 2 when the non-selectivity profiles of two MSs in comparison, are similar. The average of ζ is of course expected to be close to 1 for all study designs. However, skewness and kurtosis may potentially take off due to outlier proneness of ζ based on Deming regression.

The same sequence of percentiles as before of ζ is of interest as well as the fourth first moments based the raw simulation results. We observe that smaller study designs yield the smallest lower and most prominent upper quantiles compared to larger study designs. However, the majority (e.g., 95%) of zeta values are smaller than 2 for all relevant study designs. Smaller study designs are also more prone to outliers than larger study designs. We conclude this by examining the top graph in figure 3, where the probability of having more extreme values than 2 is approximately 2.25% for the minor study design and approximately 1.30% for the most prominent study design. The first four moments of zeta are as expected with the proneness of outliers in mind. Kurtosis is extremely large because of the outliers, and skewness is also affected. The mean and variance are also easily dominated by outliers, which is why the variance and mean values are more significant than what we saw for simulation setting 1. Another interesting note is that the distribution of ζ follows a log-normal distribution quite closely if we find an appropriate value for the second parameter. However, the difference lies in the number of outliers the distribution of ζ produces. The log-normal distribution produces very few outliers meaning that its kurtosis is much smaller than what we see for theoretical distribution of ζ.

Simulation results for the third set of simulations

In the third set of simulations, we will consider the same study designs as we did for the second set of simulations. However, we will implement concentration dependent MS SDs in addition. The simulated ζ values will be simulated for relationships between MS SDs and concentration considering both non-decreasing ($\{\eta \geq 1, \eta_0 = 1\}$) and decreasing ($\{\eta < 1, \eta_0 = 1\}$) relationships. Heteroscedasticity defined in this was are unlikely to affect the mean of ζ. Nevertheless, variance, kurtosis and skewness may potentially increase.

From figure 5, we observe an obvious relationship between η and ζ. All moments shows an increasing pattern referring to the truncated ζ values. The percentage increase from smallest to largest heteroscedasticity factor in terms of moments are:

- mean: 14 %
- variance: 3429 %
- skewness: 109 %
- kurtosis: 168 %

The illustration in figure 1 illustrates the largest and smallest heteroscedasticity factors. Potential impact on ζ caused by heteroscedasticity are related to the magnitude of the heteroscedasticity factor. For mean and variance, this relationship appears to be exponential, where the slope of the variance vs. η relationship are considerably steeper than the curve for mean vs. η relationship. The variance is accordingly expected to take off even for smaller values of η. However, the mean will be much more stable, as the percentage increase between largest and smallest η values are only 14 %. Interestingly, skewness and kurtosis seem to increase steeply at first, but then start flatting out for η values beyond 2. Based on figure 6, the empirical quantiles of raw ζ values and truncated ζ values overlaps quite well up to $\zeta = 2$, but from there raw ζ values propose a heavier tail than the truncated ζ values. Heavier tails is of course expected due to the large difference in skewness and kurtosis.

Simulation results for the fourth set of simulations

In the fourth set of simulations, we will consider the same study designs as before, but now with differences in non-selectivity defined by simulation setting 3, that is, random differences in non-selectivity. Larger values of p and $m_{\max}$ are expected to result in enlarged means of ζ. However, we also expect random differences in non-selectivity to impact variance, skewness and kurtosis.

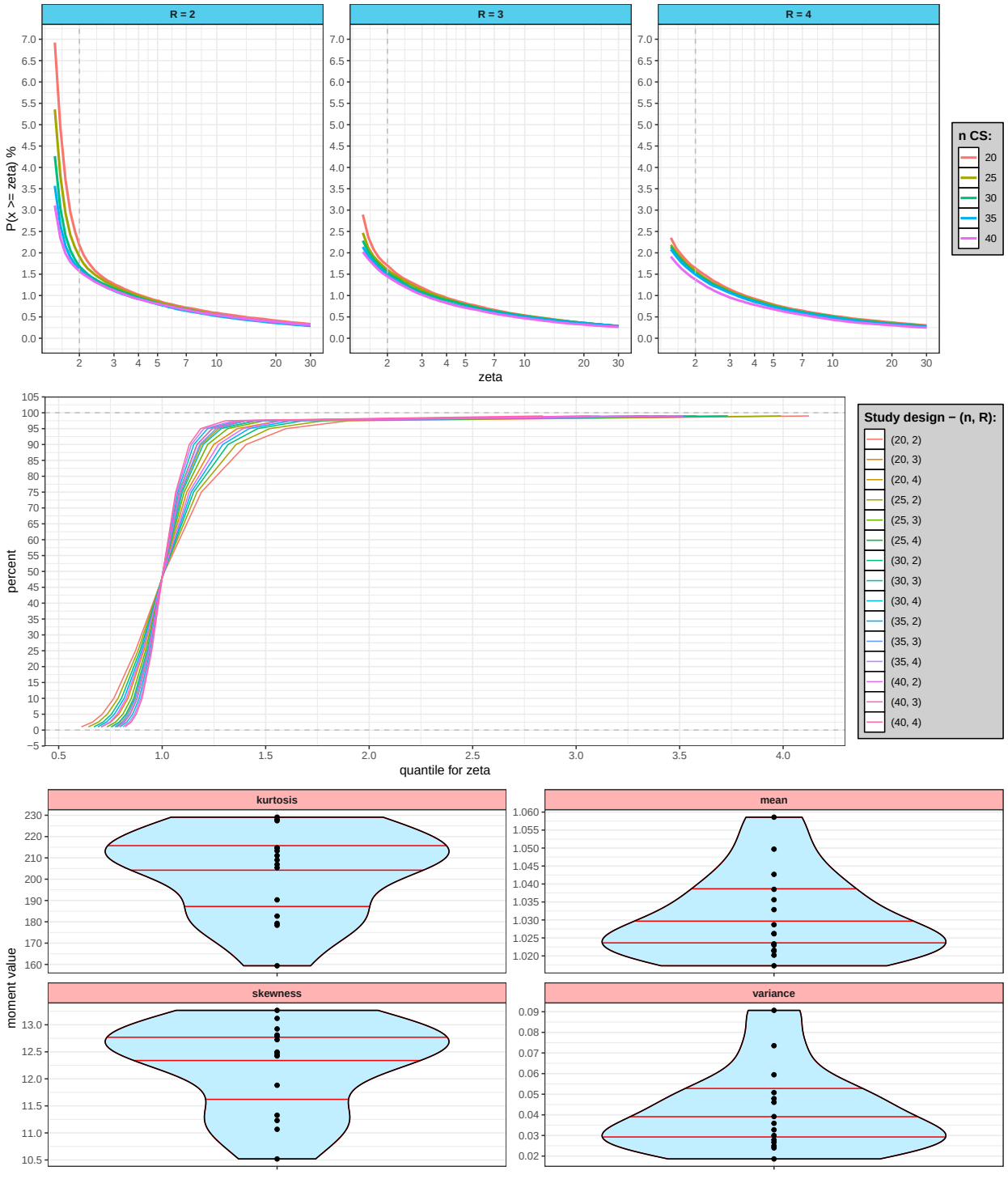


Figure 4: Simulations results for zeta based on different study design when there are no differences in non-selectivity

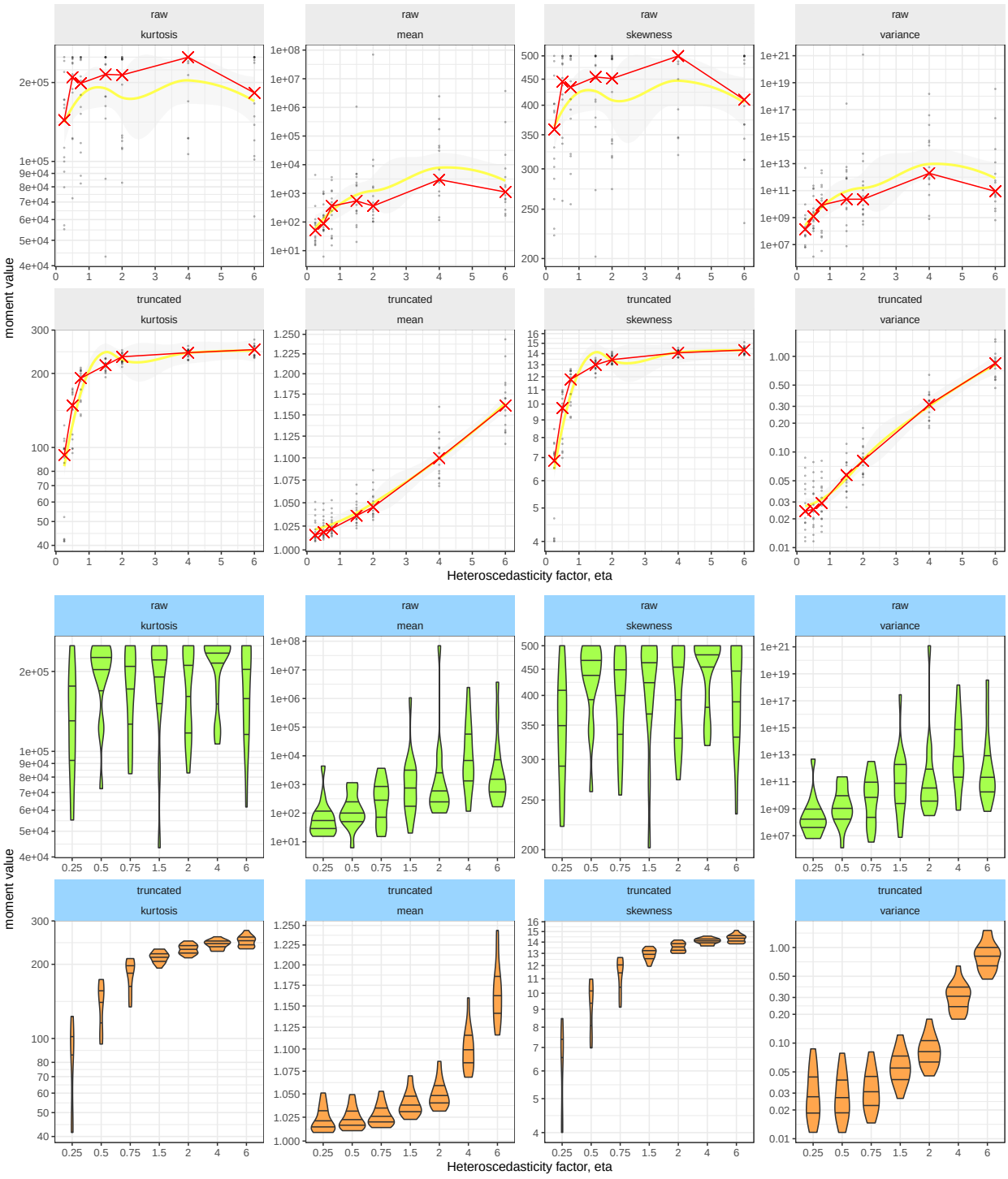


Figure 5: Relationship between heteroscedasticity factor and study study design, and moments of zeta.

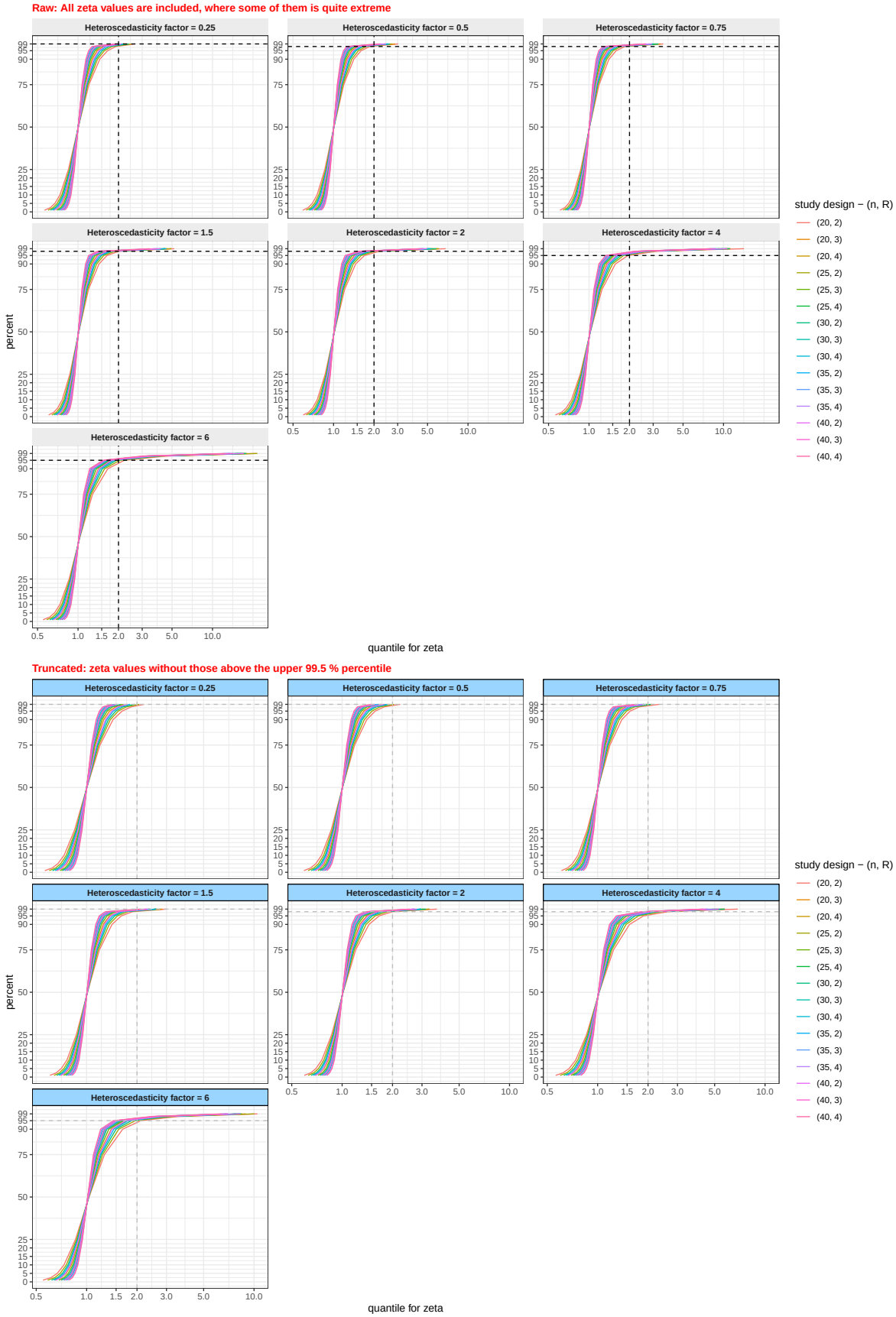


Figure 6: Relationship between study design heteroscedasticity factor and zeta's quantiles

Figure 7: Moments of zeta for all combinations of simulation parameters concerning the random dins setting

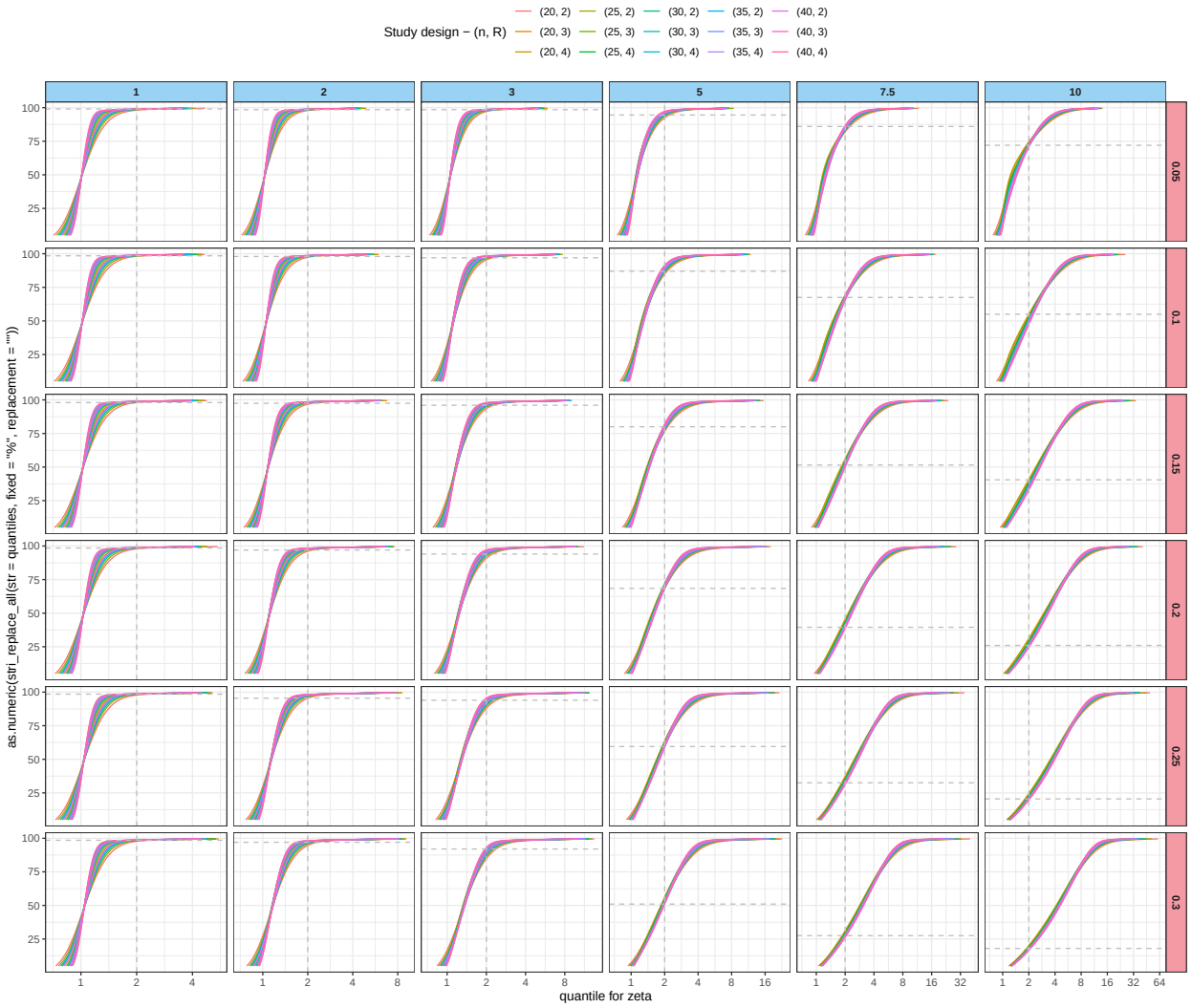
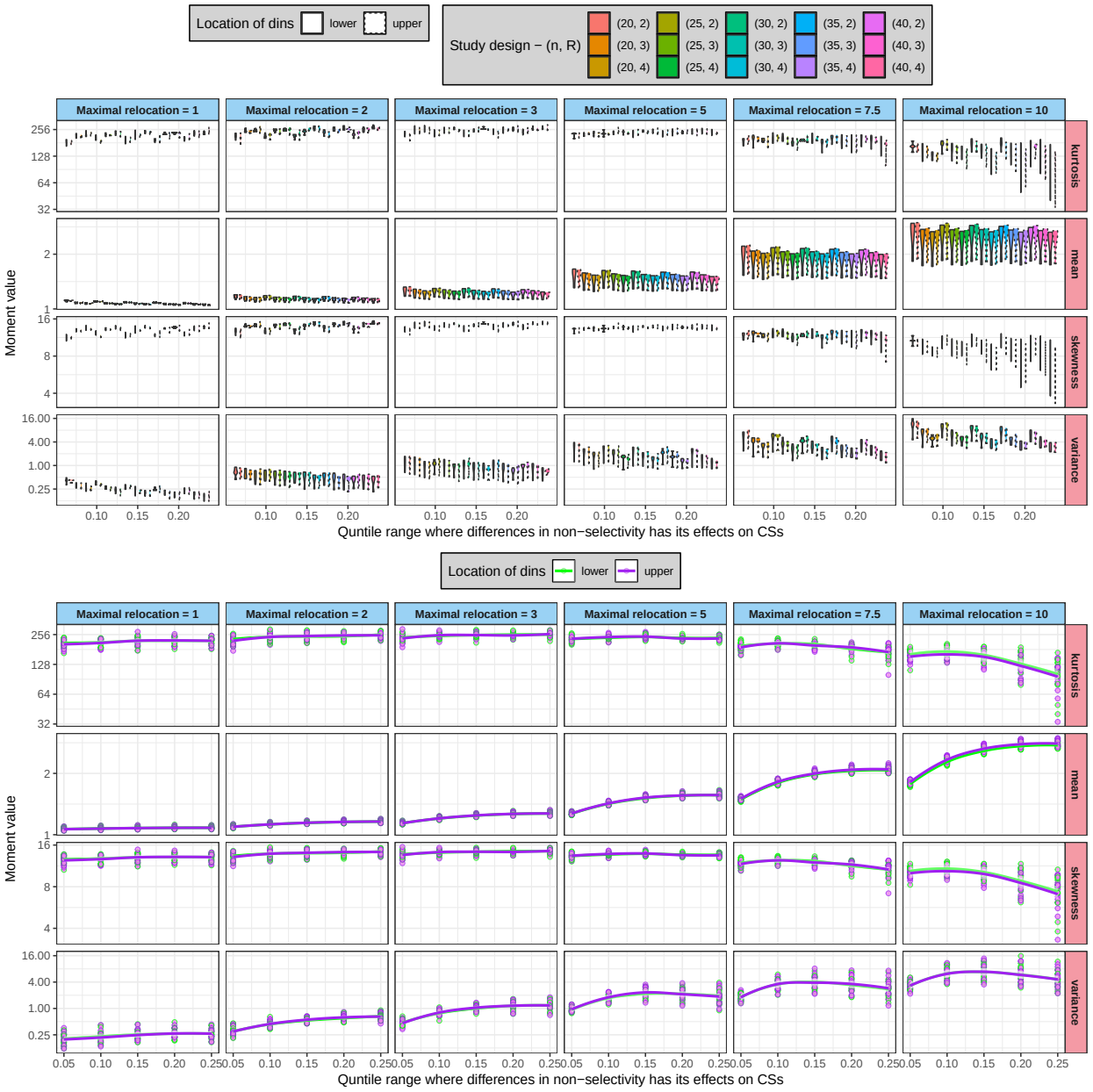
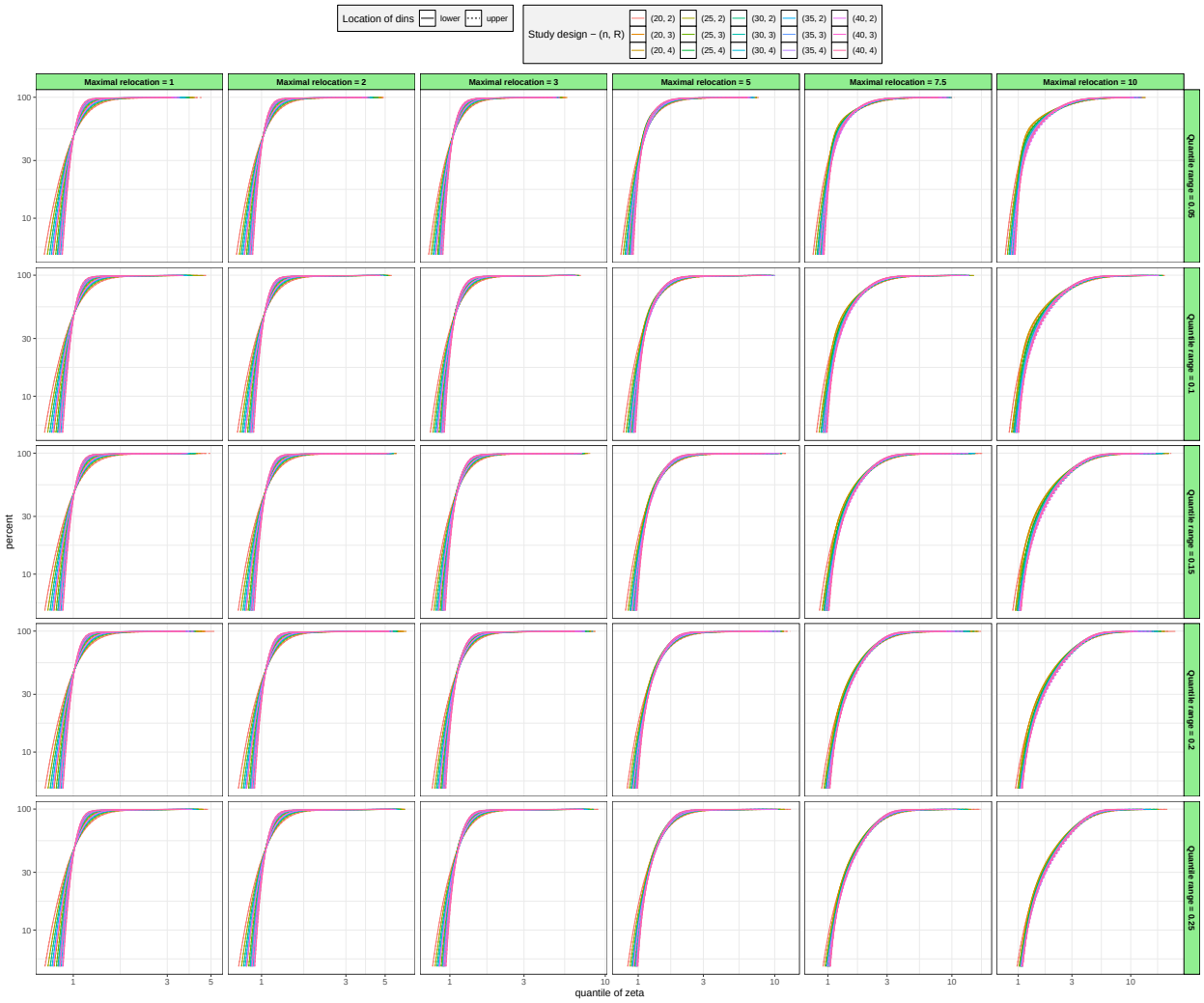


Figure 8: Quantiles for all combination of simulation parameters regarding the random dins setting


Location of dins
lower
upper
Study design − (n, R)
(20, 2)
(20, 3)
(20, 4)
(25, 2)
(25, 3)
(25, 4)
(30, 2)
(30, 3)
(30, 4)
(35, 2)
(35, 3)
(35, 4)
(40, 2)
(40, 3)
(40, 4)
Maximal relocation = 1
Maximal relocation = 2
Maximal relocation = 3
Maximal relocation = 5
Maximal relocation = 7.5
Maximal relocation = 10
kurtosis
mean
skewness
variance
Moment value
Quntile range where differences in non−selectivity has its effects on CSs


Location of dins
lower
upper
Study design − (n, R)
(20, 2)
(20, 3)
(20, 4)
(25, 2)
(25, 3)
(25, 4)
(30, 2)
(30, 3)
(30, 4)
(35, 2)
(35, 3)
(35, 4)
(40, 2)
(40, 3)
(40, 4)
Maximal relocation = 1
Maximal relocation = 2
Maximal relocation = 3
Maximal relocation = 5
Maximal relocation = 7.5
Maximal relocation = 10
Quantile range = 0.05
Quantile range = 0.1
Quantile range = 0.15
Quantile range = 0.2
Quantile range = 0.25
percent
100
30
10
1
3
5
10
quantile of zeta