

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

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

August 26, 2022

Introduction

There is often a discussion about the meaning of selectivity and what it includes, especially in analytical chemistry. In that context, there is also a discussion whether it is a qualitative term or whether it can be quantified (1). In VIM, the definition of selectivity is “*Selectivity of an MP is a property whereby the measured value of a measurand is independent of other measurands or other quantities in the sample (ref VIM). Other measurands or quantities may be metabolites of the measurand or molecular forms of the measurand, other ions or molecules, or influences on the measurement from any source other than the measurand itself.*” Selectivity of a measurement system, MS, was formerly called *analytical specificity*. In the description of selectivity in a Eurochem guide (2), interferences such as hemolysis and lipidemia are implicitly included in the definition. The Eurochem Guide also gives practical examples of how selectivity can be examined. Evaluation of selectivity is an essential part of a harmonization protocol as well as in method comparisons, especially when a method is compared to a reference measurement procedure (3, 4). MSs with minimal non-selectivity or at least similar non-selectivity should ideally be used. Accordingly, it is vital to quantify relative differences in non-selectivity by employing a quantitative approach. Differences in non-selectivity refer to whether two compared MSs are similar or not. Suppose one of the MSs is considered selective. In that case, any non-selective MS will enable us to detect non-selective MS by comparing the selective MS to candidate MSs. In principle, an estimate of differences in non-selectivity can be calculated as the ratio of the estimated average variance of a regression model’s residuals – to the expected theoretical minimal variance of the residuals estimated by variances of the MSs in the comparison. In case of no differences in non-selectivity, this ratio is expected to be close to one.

Methods

Theory

Derivation of the difference in non-selectivity estimator

We seek a criterion for detecting the effects of excessive relative non-selectivity on a comparison between a pair of MSs, that we will refer to by MS_X and MS_Y . To obtain a meaningful criterion of difference in non-selectivity between compared MSs, we need to consider the principle of differences in non-selectivity from a mathematical perspective. If excessive relative differences in non-selectivity can be formulated mathematically, it may be straight-forward to determine an appropriate rule for detection of excessive differences in non-selectivity based on statistical principles. In this text, we mathematically define excessive differences in non-selectivity as a property of two MSs in a comparison, where their clinical sample (CS) measurement results deviates more than we can expect by random variation caused by analytical uncertainty (e.g., variance) of the compared MSs. We define $SD_{MS_X}^2$ to be the estimated variance of MS with measurement results in the x -direction and $SD_{MS_Y}^2$ to be the estimated variance of the MS with measurement results in the y -direction. Before we mathematically define $SD_{MS_X}^2$ and $SD_{MS_Y}^2$, we denote individual measurement results from MS_X by x_{ir} . The

corresponding measurements from MS_Y is denoted by y_{ir} . The indices i and r signify CS id and replicate measurement id, respectively. For example, $i = 3$ and $r = 2$ correspond to the second replicate measurement on the CS with id equal to 3. It is well established that $SD_{MS_X}^2$ and $SD_{MS_Y}^2$ may be calculated by the pooled within-CS variances over all unique CSs (ref):

$$SD_{MS_X}^2 = \frac{1}{n_X} \sum_{i=1}^{n_X} \frac{1}{R_i - 1} \sum_{r=1}^{R_i} (x_{ir} - \bar{x}_i)^2$$

$$SD_{MS_Y}^2 = \frac{1}{n_Y} \sum_{i=1}^{n_Y} \frac{1}{R_i - 1} \sum_{r=1}^{R_i} (y_{ir} - \bar{y}_i)^2$$

Dividing $SD_{MS_Y}^2$ by $SD_{MS_X}^2$ yields the estimate of Λ , which is the theoretical ratio of MS variances. We will refer to the estimate of Λ by λ . We will define a measure of differences in non-selectivity between MSs that depend on the variance of a regression model's residuals and the sum of $SD_{MS_Y}^2$ and $SD_{MS_X}^2$. The reasoning behind this principle is the bias-variance trade-off in principle typically used in statistical learning (ref). The bias-variance trade-off principle states that the expected squared prediction error of a new observation, which in our case is a new measurement result, may be decomposed into *statistical squared model bias*, *model variance* and *irreducible variance*. The magnitude of the two former quantities of the three may be practically viewed as the amount of differences in non-selectivity between the compared MSs. The irreducible variance can namely be further decomposed into

$$\text{irreducible variance} = \sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2 \quad (1)$$

where $\sigma_{MS_X}^2$ and $\sigma_{MS_Y}^2$ are the theoretical MS variances that $SD_{MS_X}^2$ and $SD_{MS_Y}^2$ are estimating. Moreover, β_1 is the regression slope coefficient, that may be estimated based on the chosen regression model. We accordingly define the theoretical quantity based on our principle for difference in non-selectivity by ζ^* (zeta star) in terms of variability caused by differences in non-selectivity, D :

$$\zeta^* = \frac{E[(y_0 - \hat{y}_0)^2]}{\sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2} = \frac{D + \sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2}{\sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2} \quad (2)$$

Here, E is the expectation operator, y_0 is a new measurement based on MS_Y , and \hat{y}_0 is the predicted measurement based on MS_Y using the new measurement x_0 based on MS_X . Suppose we assume that our regression model is unbiased, which is equivalent to

$$E[y_0 - \hat{y}_0] = 0,$$

then we can redefine ζ^* :

$$\zeta^* = \frac{\text{Var}[y_0 - \hat{y}_0]}{\sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2} = \frac{D + \sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2}{\sigma_{MS_Y}^2 + \beta_1^2 \sigma_{MS_X}^2} \quad (3)$$

Using Deming regression based on (Fuller, Gillard), $\text{Var}[y_0 - \hat{y}_0]$ may be estimated by

$$\widehat{\text{Var}}[y_0 - \hat{y}_0] = \widehat{\text{Var}}[b_1] s_{XX} + \widehat{\text{Var}}[b_1] \widehat{\sigma_{MS_X}^2} + (1 + n^{-1} R^{-1})(\beta_1^2 + \lambda) \widehat{\sigma_{MS_X}^2}, \quad (4)$$

assuming that $R_i = R \forall i \leq n$. See appendix for details on derivation. Here, s_{XX} is the maximum likelihood estimator for the variance of measurements from MS_X , $\widehat{\text{Var}}[b_1]$ is the asymptotically estimated variance of the Deming slope estimator b_1 . Moreover, $\widehat{\sigma_{MS_X}^2}$ is another estimator $\sigma_{MS_X}^2$, that unlike $SD_{MS_X}^2$ accounts for the relationship between MS_X and MS_Y . See (refs) for expressions for b_1 , $\widehat{\text{Var}}[b_1]$ and $\widehat{\sigma_{MS_X}^2}$. Lastly, R is the common total number of replicated measurements performed on each CS, and n is again the number of unique CSs. As we learn in the supplemental file, the Deming approach to estimate $\text{Var}[y_0 - \hat{y}_0]$ shows are prone to estimate wild extreme values of ζ . Actually, depending on the study design, 0.5 to 6 percent of all simulated ζ values will be such wild values. Another, more pragmatic estimator for $\text{Var}[y_0 - \hat{y}_0]$ may be found by ordinary least squares regression:

$$\widehat{\text{Var}}[y_0 - \hat{y}_0] = \frac{\sum_{i=1}^n \sum_{r=1}^R (y_{ir} - \hat{y}_{ir})^2}{nR - 2} \cdot \frac{nR + 2}{nR}, \quad (5)$$

The quantity

$$\frac{\sum_{i=1}^n \sum_{r=1}^R (y_{ir} - \hat{y}_{ir})^2}{nR - 2},$$

is recognized to be the estimator of the irreducible variance, S^2 , of the ordinary least square model, that also captures D . Accordingly, we have the closed-form plug-in estimator of ζ^* :

$$\zeta = \frac{S^2 \cdot \frac{nR+2}{nR}}{SD_{MS_Y}^2 + b_1^2 SD_{MS_X}^2} \quad (6)$$

For brevity, the notation $S_{P_{AR}}^2 = S^2 \cdot \frac{nR+2}{nR}$ is introduced, providing our difference in non-selectivity measure estimate ζ . P stands for prediction error, and AR stands for all replicates, signifying that all replicates are used to estimate the ordinary least squares regression model. Thus,

$$\zeta = \frac{S_{P_{AR}}^2}{SD_{MS_Y}^2 + b_1^2 SD_{MS_X}^2}. \quad (7)$$

Statistical properties of the difference in non-selectivity estimator

Based on (equation), ζ must be a random variable, because it is defined by estimators that per definition are random. As part of determining which values of ζ that should alarm us, we are required to examine the statistical distribution of ζ based on numerous sets of statistical assumptions. For example, we expect that the distribution of ζ differs between study designs. Based on a set of distributional assumptions on the CSs' measurements, such as normality, it is possible to formulate the difference in non-selectivity principle in terms of a hypothesis test, by deriving that ζ e.g., follows a F-distribution asymptotically under certain conditions. Nevertheless, in practice, our study designs are somewhat restricted, so this would not be an optimal solution. However, assuming our random samples of CSs measurements to represent the theoretical population of CS measurements suitably we may employ Monte Carlo simulation based on non-parametric resampling to do inferences on ζ . Monte Carlo simulation based on non-parametric resampling is a method typically denoted by *bootstrap*. Thus, only requiring a representable random sample of the theoretical population of interest, enable us to bypass distributional presumptions, and still being able to do inference on ζ . Using bootstrap will provide us with different types of bootstrap confidence intervals (e.g., bootstrap percentile confidence interval), estimation of standard error, and more concerning ζ . In our case, we are required to use a cluster-bootstrap algorithm, because each of our CSs are typically measured in replicate, meaning that each CS is a cluster. As (Afron and Hastings ref) states, we are only required to resample the clusters (CSs), and consequently no hierarchical resampling. Suppose that B resamples of ζ is required. Then the cluster bootstrap algorithm is:

For $b = 1, \dots, B$:

1. Draw n CSs with replacement from original data, that gives resampled data.
2. Calculate ζ_b based on the b th resampled data

Based on the set of B resampled ζ values, that is $\zeta_1 \dots, \zeta_B$ and the original ζ , we can for example estimate bias and variance of ζ :

$$\widehat{\text{bias}}[\zeta] = \frac{1}{B} \sum_{b=1}^B (\zeta - \bar{\zeta}_b)$$

$$\widehat{\text{Var}}[\zeta] = \frac{1}{B-1} \sum_{b=1}^B (\zeta_b - \bar{\zeta}_b)^2$$

Defining the order statistic of $\zeta_1 \dots, \zeta_B$ by $\zeta_{(1)} \dots, \zeta_{(B)}$, we can also obtain bootstrap estimates of percentiles of ζ . For example, if $B = 1000$, the bootstrap estimates the median, lower quartile and upper quartile of ζ

are respectively,

$$\begin{aligned}\widehat{\text{median}}[\zeta] &= \zeta_{(500)} \\ \widehat{\zeta}_{0.25} &= \zeta_{(250)} \\ \widehat{\zeta}_{0.75} &= \zeta_{(750)}\end{aligned}$$

In the same example, we would also estimate the 95% percentile bootstrap confidence interval for ζ :

$$\widehat{\text{CI}}_{\zeta} = [\zeta_{(25)}, \zeta_{(975)}]$$

Investigation of the difference in non-selectivity estimator using Monte Carlo simulations

We have formerly discussed the theoretical and general relationship between ζ and potential differences in non-selectivity between compared MSs. However, we are now interested in monitoring the the distribution of ζ with various degrees of differences in non-selectivity between pairs of MSs. The distribution of ζ will be estimated using simple parametric Monte Carlo simulations for various underlying theoretical populations to do a comprehensive investigation of ζ and determine which ζ values that should represent excessive differences in non-selectivity. In this section, we will refer to six main simulation situations, wherein each ζ is examined for various combinations of simulation parameters such as MS CVs, n , R and relevant parameters for the simulation situation in question. These six sets of simulations making each simulation situation are further outlined in the supplemental file. The summary of the simulation situations are the following:

For each simulation situation we simulate at least 100 000 ζ values based on each combination of common simulation parameters, that are n , R , MS CVs and concentration intervals. After concluding that MS CVs and concentration intervals are pragmatically negligible, which is done in the supplemental file, we will only consider all combinations of $n = 20, 25, 30, 35, 40$ and $R = 2, 3, 4$ (fifteen unique combinations), and let MS CVs and concentration intervals be drawn from random variables for each simulated ζ . See details regarding the distribution of MS CVs and concentration intervals in supplemental file. In addition to these base simulation parameters, we will include other parameters such as heteroscedasticity factor and systematic differences in non-selectivity magnitude.

1. All MS pairs have identical non-selectivity profiles
2. All MS pairs have identical non-selectivity profiles, but heteroscedasticity is introduced
3. All MS pairs have non-zero differences in non-selectivity profiles defined by differences in non-selectivity setting 1
4. All MS pairs have non-zero differences in non-selectivity profiles defined by differences in non-selectivity setting 2
5. All MS pairs have non-zero differences in non-selectivity profiles defined by differences in non-selectivity setting 3
6. All MS pairs have non-zero differences in non-selectivity profiles defined by differences in non-selectivity setting 3, but in combination with heteroscedasticity.

All simulations will be run by R wherein the C++ package [fasteqa](#) is used. The simulation results are presented in the figures of the supplemental file and made reproducible with a seed for those wanting to re-run them.

Results

When two MSs have identical or near identical non-selectivity profiles, ζ is typically expected to be in between 0.5 and 1.5. Of course, larger study designs decreases the range of possible ζ values. However, given an appropriate study design ζ is improbable to be either below 0.5 or above 1.5. As expected the mean ζ values are between 1.0 and 1.06 for MSs with no differences in non-selectivity. The CV of ζ is between 13.34% and 29.41%. Furthermore, skewness is positive which means that the distribution of ζ is right-skewed with empirical skewness estimates between 10 and 14. Kurtosis, which measures outlier tendencies, decreases slightly as the number of CSs increases, but much more when including more replicated measurements. So in

order to decrease kurtosis, one require more replicates to be performed on the CSs. If a ζ value is larger than e.g., 2, we are very certain that there is differences in non-selectivity between the compared MSs. However, we will later see that a small ζ value, e.g., < 1 does not necessarily imply that there are no differences in non-selectivity.

If two MSs have identical or near identical non-selectivity profiles, but with heteroscedasticity, the distribution of ζ is altered concerning the uncertainty (variance), skewness and kurtosis. However, the effect on the average of ζ is negligible. A potential fix would be to account for heteroscedasticity in the linear model proposed by the two MSs in comparison, so that heteroscedasticity have less drastic impact on potential conclusions drawn. Interestingly, as the heteroscedasticity increases in magnitude, kurtosis and skewness appear to level out. However, this is not the case for the variance which is quickly increasing independently of the heteroscedasticity magnitude.

In case of non-zero differences in non-selectivity, that has a random impact on the CSs, we observe that the relocation magnitude multiplier is much more influential than the number of CSs affected by differences in non-selectivity. As the relocation magnitude multiplier increases from 0 to 10 the average value of ζ increases from 1 to a number between 3 and 4. The variance is increasing drastically from 0.02-0.09 to 5-30. This increase in variance makes things difficult, because ζ values of < 0.5 is more probable than before. This entails that if we get a small ζ values (e.g., below 1.5) we cannot be sure that it actually represents no differences in non-selectivity. The standard deviation or variance of ζ , or a confidence of ζ should always be included, so that we make more robust conclusions regarding differences in non-selectivity.

Discussion

Appendix