Small Sample Size Capability Index for Assessing Validity of Analytical Methods

E. Rozet[1], B. Boulanger [2], E. Ziemons [1], R.D. Marini [1], Ph. Hubert [1]
Eric.Rozet@ulg.ac.be
[1] Analytical Chemistry Lab., University of Liège, Liège, Belgium.
[2] Arlenda s.a., Liège, Belgium.

INTRODUCTION & AIM

- The commonly used formulas to compute capability indices such as Cpk, will highly overestimate the true capability of analytical methods. Especially during methods validation or transfer, where there are only few experiments performed and, using in these situations the commonly applied capability indices to declare a method as valid or as transferable to a receiving laboratory will conduct to inadequate decisions.
- In this work, an improved capability index, namely Cpk-tol is proposed. Through Monte-Carlo simulations, they have been shown to greatly increase the estimation of analytical methods capability in particular in low sample size situations as encountered during methods validation or transfer.

MODIFIED CAPABILITY INDEX

The core problem when using capability indices in validation studies is the lack of sufficient data to estimate precisely the mean and standard deviation of the analytical method. In order to circumvent this and to take into account the uncertainty of the analytical method mean and standard deviation when computing Cpk, the use of tolerance interval should be preferred.

Additionally, the estimation of the mean and standard deviation of analytical methods should be made following the statistical model representing the way experiments have been performed. Method validation experiments or method transfer experiments are following a hierarchical or stratified sampling scheme that should be taken into account when computing analytical mean results and standard deviation and therefore to compute capability indices. For these stratified random sampling schemes commonly encountered during methods validations or transfers, a β-expectation tolerance intervals formula is given by Mee [1]:

$$[L,U] = [\hat{\mu} - k_E \hat{\sigma}_{IP}; \hat{\mu} + k_E \hat{\sigma}_{IP}] \qquad \text{(Eq.1)}$$

$$\text{where } k_E = t_{(df,(1+\beta)/2)} \sqrt{1 + \frac{J\hat{R}+1}{JI(\hat{R}+1)}}$$

$$\text{with } df = \frac{(\hat{R}+1)^2}{\frac{\left(\hat{R}+\frac{1}{I}\right)^2}{J-1} + \frac{\left(1-\frac{1}{I}\right)}{JI}} \quad \text{and} \quad \hat{R} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2}$$

t(df , γ) is the γ[th] percentile of a Student distribution with df degrees of freedom and $\hat{\mu}$ is the estimated mean of the results. The intermediate precision variance can be estimated using: $\hat{\sigma}_{IP}^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2$. $\hat{\sigma}_B^2$ is the run-to-run or series-to-series variance and $\hat{\sigma}_W^2$ is the within-run or repeatability variance obtained with random one way ANalysis Of Variance (ANOVA) methodology [2]. J is the number of series performed and I the number of replicates per series.

The modified capability index proposed, Cpk-tol, is thus based on these tolerance intervals and is computed as it follows:

$$Cpk - tol = \min\left[\frac{USL - \hat{\mu}}{t_{(df,(1+\beta)/2)}\sqrt{1+\frac{J\hat{R}+1}{JI(\hat{R}+1)}}\hat{\sigma}_{IP}}, \frac{\hat{\mu} - LSL}{t_{(df,(1+\beta)/2)}\sqrt{1+\frac{J\hat{R}+1}{JI(\hat{R}+1)}}\hat{\sigma}_{IP}}\right] \qquad (Eq.2)$$

The Cpk index is computed with 3 at the denominator meaning that for a centred process the maximum fraction of non conforming result is about 2,700 dpm (precisely 2,699.796 dpm). In order to keep this same theoretical coverage of the distribution used with the Cpk index (i.e. $\pm 3\sigma$), the probability β of the Cpk-tol index is fixed to 0.9973.

SIMULATIONS

Independent validation results were generated from the random one-way ANOVA model described below:

$$X_{ij} = \delta + \phi_B + \varepsilon_W \quad \text{Eq. 3.}$$

where $X_{ij}$is the result of the j[th] measurement in series i, $\delta = \mu_{Lab} - \mu_T$ is the bias between the true (or reference or nominal) value of the result (μ_T) and the average value of the results of the laboratory (μ_{Lab}), $\phi_B$is the between series random effect supposed to be normally distributed $N(0,\sigma_B^2)$ and ε_W is the within-series (or repeatability) random error supposed to be independent and normally distributed $N(0,\sigma_W^2)$.

SIMULATIONS RESULTS

From the Figures 1a to 1f, it can be seen that the probability to exceed the true capability value (defined by the triangle in continuous line of Figures 1a-1f) by using the Cpk index is almost 50%, whatever the true capability value and whatever the sample size used in the method validation. Indeed the isoprobabilty curve (dashed line of Figures 1a-1f) that is almost exactly on the region which defines methods with known true Cpk value is the isoprobability curve of 50%.

By opposition when using Cpk-tol, Figures 1a to 1f show that the probability to exceed the true capability value is extremely low as the closest isoprobablity curve to the region defining methods with true capability indices of 1, 1.33 or 2 is the 10% one. Therefore there is only about 10% probability to declare a method capable when in reality it is not, i.e. the costumer risk is about 10% using such a capability index compared to the 50% risk observed for the classical Cpk index.

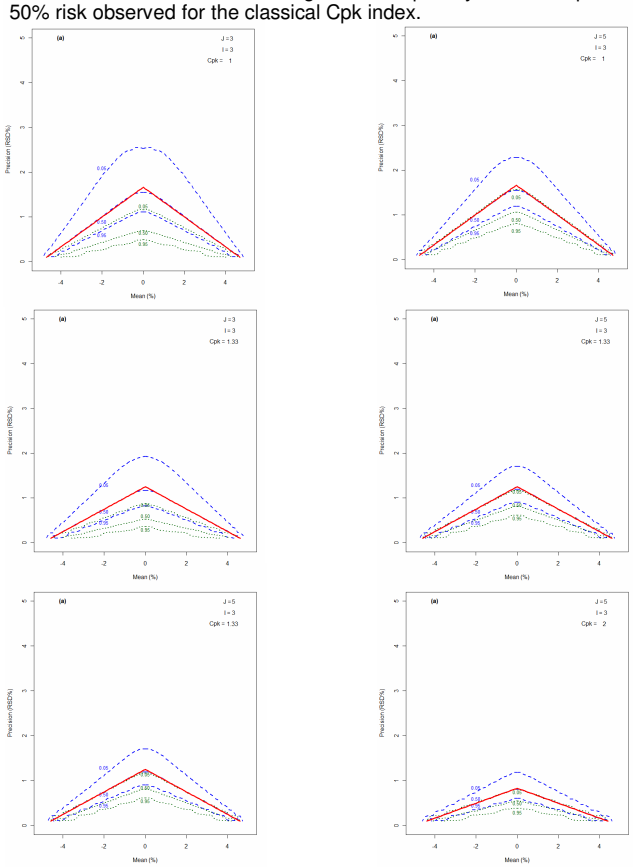

Fig. 1. Isoprobability contour measuring the probability that Cpk (dashed curves) and Cpk-tol (dotted curves) exceeds the true Cpk value of (a) 1, (c) 1.33 and (e) 2 with a design of 3 series and 3 repetitions per series or exceeds the true Cpk value of (b) 1, (d) 1.33 and (f) 2 with a design of 5 series and 3 repetitions per series.

While the Cpk-tol index controls well the costumer risk, the producer risk (i.e. the risk to conclude the method is not capable while it is truly capable) is relatively high. However this risk can be reduced by increasing the sample size of the method validation. This is shown by comparing Figures 1, 3, 5 obtained with 3 runs and 3 repetitions per run to Figures 2, 4 and 6 obtained with 5 runs and 3 repetitions per run for true Cpk values of 1, 1.33 and 2, respectively.

CONCLUSIONS

Finally, these simulations highlighted first the fact that using Cpk index to decide about the validity of analytical methods is highly controversial especially when using a method validation design of 3 runs and 3 replicates. Second, they showed that using Cpk-tol to make such a decision better controls the costumer risk, thus controls the risk for patients or public health risk, while the producer risk can be modulated by increasing sample size.

[1]. R.W. Mee, Technometrics 26 (1984) 251.
[2]. Searle S.R., Casella. G. and McCulloch C.E., Variance components (1992), Wiley.

Contact: Eric.Rozet@ulg.ac.be or:

ACKNOWLEDGMENTS

A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is gratefully acknowledged.

Reprints: http://hdl.handle.net/2268/104084/ or: