



BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
FACULTY OF CHEMICAL AND BIOENGINEERING
OLÁH GYÖRGY DOCTORAL SCHOOL

STATISTICAL PROBLEMS IN THE
PHARMACEUTICAL ANALYSIS

Thesis booklet

Author: Balázs Bánfai

Advisor: Sándor Kemény,
D.Sc., Professor

Department of Chemical and Environmental
Process Engineering

2012

INTRODUCTION

1

Regulatory compliance situations frequently emerge in pharmaceutical quality assurance. I investigated the statistical background of such problems and provided new statistical methods for them. The following problems are discussed in my dissertation and statistically sound procedures are provided based on interval hypotheses. In analytical method transfer (I) it has to be proved that a method has the same performance in the applying laboratory than in the developing laboratory. During method validation the accuracy (II) of analytical methods has to be examined, to prove that the added (standard) equals the recovered (calculated) concentration. The single-point calibration (III) is widely used because of its simplicity, when the concentration–signal function is linear with zero intercept. This has to be checked during method validation as well; a decision is made about the future applicability of this simplification. The content uniformity (IV) assessment reflects the homogeneity of a product batch: if the majority of the tablets contain the active ingredient with a concentration sufficiently close to the label claim (nominal content).

LITERATURE REVIEW

2

Statistical significance tests are usually carried out by applying hypotheses of no difference, both in the pharmaceutical industry and in other industrial or scientific fields. Null hypotheses of no difference have a drawback: increasing the variance or smaller sample size leads to the acceptance of no difference with a higher probability (larger and larger deviations prove to be significant). An other problem is that accepting the null hypothesis does not imply that it is true, and irrelevant differences can become significant with small variance. Because of these problems, hypotheses of no difference are not valid for proving regulatory compliance. The two one-sided t -tests (TOST) has been known for bioequivalence problems for decades¹, but it has not dispersed into other disciplines. It is mentioned as an option² for the evaluation of analytical method

1 Schuirmann D.J. (1987) A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Pharmacodynamics* 15(6): 657–680. doi:10.1007/BF01068419

2 Ermer J. and Miller J.H.M. (2005) *Method Validation in Pharmaceutical Analysis. A Guide to Best Practice*. 1st edn. Wiley-VCH, Weinheim, Germany. ISBN: 978-3-527-31255-9 pp. 63–79.

transfer (I), but it is not required by the pharmacopeias³ or FDA guidelines, even if it has clear advantages compared to the two-sample t -test. The interval hypotheses (similar to the TOST) can be extended to regression as well.

In the accuracy (II) assessment the task is to prove that equality of the added and recovered (calculated) concentration. The USP requirements⁴ state that the lack of significance for a test on the slope being 1 is not an adequate solution (the same problems exist as in the (I) case). However, there is no proper solution given in the text. The general regression test (a simultaneous test for the slope and intercept) is a well-known method available in regression textbooks, but it does not solve the above mentioned problems. My proposed method is new in the literature.

The evaluation of single-point calibration (III) is performed during method validation, where a decision has to be made on the applicability of this simplification in routine use. The usual decision is based on a hypothesis test on zero intercept, but a regulatory requirement does not exist. This test involves the same problems as in case (I). The developed method is new in the literature as well.

The content uniformity (IV) reflects the homogeneity of the active ingredient among the tablets of a product batch. The current regulations are based on the modification of the European Pharmacopoeia (Ph. Eur.)⁵ from 2007, which were adopted by the other pharmacopeias. This required a criteria based on tolerance intervals. A tolerance interval is an interval in which a given proportion (content) of the population falls with a certain probability (confidence). The tolerance factor from the current criteria (which reflects the content and the confidence) is too small, causing a too narrow tolerance interval (and easy acceptance). This cannot be justified by practical or statistical reasons. An other problem is that the two sources of variance of a content uniformity (CU) measurement (inhomogeneity and analytical error) is overlooked.

3 CALCULATION METHODS

To compare the two-sample t -test and the two one-sided t -tests (TOST) the acceptance probabilities are calculated by integrating the joint probability den-

³ USP (2011) *General Chapters: (1224) Transfer of Analytical Procedures* in: United States Pharmacopeia—National Formulary (USP 35–NF 30). The United States Pharmacopeial Convention, Rockville, MD, USA. ISBN: 1-889788-94-4

⁴ USP (2011) *General Chapters: (1225) Validation of compendial procedures* in: United States Pharmacopeia—National Formulary (USP 34–NF 29). The United States Pharmacopeial Convention, Rockville, MD, USA. ISBN: 1-889788-94-4

⁵ Ph. Eur. (2011) *General Chapter 2.9.40*. in: European Pharmacopoeia, 7th Edition, Supplement 7.2. Council of Europe, Strasbourg, France.

sity function of a normal and a χ^2 distributed variable, exploiting the fact that the \bar{x} and the s are independent random variables:

$$P_{a(tt)} = \int_0^{\infty} \left\{ \Phi \left[-\frac{(\mu_1 - \mu_2)}{\sigma} \frac{1}{\sqrt{2/n}} + t_{\alpha}/2 \sqrt{\frac{V}{2n-2}} \right] - \Phi \left[-\frac{(\mu_1 - \mu_2)}{\sigma} \frac{1}{\sqrt{2/n}} - t_{\alpha}/2 \sqrt{\frac{V}{2n-2}} \right] \right\} f_{\chi^2(2n-2)}(V) dV$$

and

$$P_{a(TOST)} = \left(\frac{\Delta}{i_{\bar{x}} \sigma} \right)^{2n(n-1)} \int_0^{\infty} \left\{ \Phi \left[\frac{-(\mu_1 - \mu_2) - \Delta}{\sigma} \frac{1}{\sqrt{2/n}} + t_{\alpha} \sqrt{\frac{V}{2n-2}} \right] - \Phi \left[\frac{-(\mu_1 - \mu_2) + \Delta}{\sigma} \frac{1}{\sqrt{2/n}} - t_{\alpha} \sqrt{\frac{V}{2n-2}} \right] \right\} f_{\chi^2(2n-2)}(V) dV$$

The decision on the applicability of single-point calibration is based on the confidence interval calculated for the bias, which is the expected value of the difference between the computed and true concentrations. Since the computed concentration is the ratio of two normally distributed random variables, and its expected value is not defined, the ratio of the two expected values will be called bias. The confidence interval can be calculated using Fieller's theorem. Rearranging the quadratic formula and substituting the parameters of the fitted line yields:

$$CI_{u,l} = (x^* - x) \frac{B \pm \sqrt{B^2 - AC}}{C},$$

where

$$A = b_0^2 - s_0^2 t^2,$$

$$B = b_0(b_0 + b_1 x^*) - (s_0^2 + s_{01} x^*) t^2,$$

$$C = (b_0 + b_1 x^*)^2 - (s_0^2 + 2s_{01} x^* + s_1^2 x^{*2}) t^2.$$

To calculate the tolerance interval for the true inhomogeneity, the one-way random factor ANOVA model can be used. The Tukey-Williams confidence bound is used for the inhomogeneity variance (s_a^2), the degrees of freedom are calculated with the Satterthwaite-approximation, and the \hat{k} tolerance factor is computed with the parameters of the random sample.

$$\bar{x} \pm \max \left\{ \hat{k} \sqrt{\max \left[0, \frac{s_a^2 - F_{(2+\gamma)/3}(v_1, v_2) s_r^2}{p} \right]}, \frac{t_{rp-1, (1-\gamma)/2}}{\sqrt{rp}} s_e \right\}$$

4 RESULTS

For the comparison of the two-sample t -test and the two one-sided t -tests the probability of accepting the method transfer (I) was computed. These probabilities depend on the variance of the methods (σ^2), the relevant difference (Δ), the type I error probability (α), and the true difference between the two methods ($\mu_1 - \mu_2$). Figure 1 shows the acceptance probabilities in function of the true difference for a selection of sample sizes (with the shown parameters).

The required sample size (n) can be computed for a relevant difference with given type II error probability (β), in this case yielding $n = 23$ (it is the same with both hypothesis tests). It can be seen that the two procedures give identical results with the required sample size. With n less than required, the manufacturer's risk (incorrect rejection) is maintained (at $\mu_1 - \mu_2 = 0$), while the consumer's risk (incorrect acceptance) depends on the sample size ($\mu_1 - \mu_2 = 1$). The properties are the opposite in case of the two one-sided t -tests, where the consumer's risk is fixed (probability of acceptance is 0.1 at $\mu_1 - \mu_2 = 1$), and the manufacturer's risk depends on the sample size.

The two one-sided t -tests (TOST) contain the practical question in the alternative hypothesis (the difference of the two laboratories does not exceed an allowed value). This guarantees that irrelevant differences can not become significant, and the acceptance of the null hypothesis proves a sufficiently small

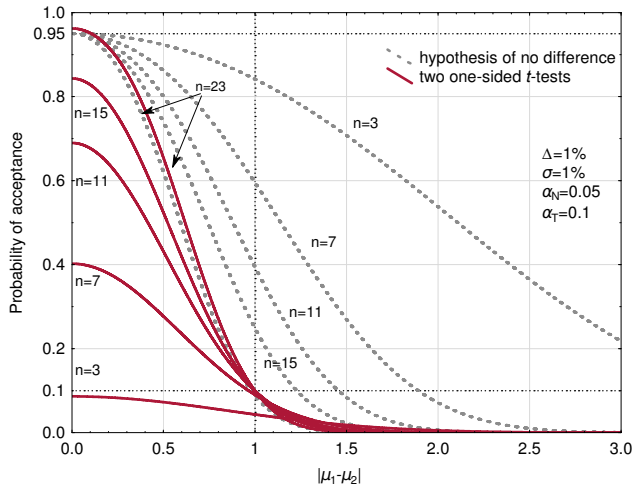


Figure 1. Probability of accepting the method transfer

difference between the laboratories. An other advantage is that the consumer's risk is maintained low, and the manufacturer's risk can be made lower with increasing sample size.

During method transfer, the actual (inhomogeneous) products have to be tested to ensure equality, so the preparation of identical samples has to be assured. This can be carried out by homogenizing a given number of tablets. I provided a calculation method to determine the required number of tablets.

In the assessment of accuracy (II) we are facing two major problems: large differences can be overlooked due to large variance, and practically irrelevant deviations can prove significant if the variance is small. An example for the latter: with small variance, performing the hypothesis tests on zero intercept and unit slope leads to significant differences. The general regression test shows a significant result too ($p = 0.037$), meaning that the accuracy is rejected. Figure 2 shows the result of my proposed method: the confidence band for the $(Y - x)/x$ recovery in function of the added concentration (x). The allowed difference is usually 2% for this measurement, and the confidence band is entirely contained in this range, so the accuracy of the method can be concluded.

If the variance is large the general regression test fails to detect the difference, but the confidence interval approach gives satisfactory results in this case as well: the confidence interval would be wider than the allowed range, so the lack of bias would be rejected.

Similar problems emerge when evaluating the applicability of the single-point calibration (III), if it is based on a hypothesis test of zero intercept. The proper procedure is the examination of the bias caused by using the single-

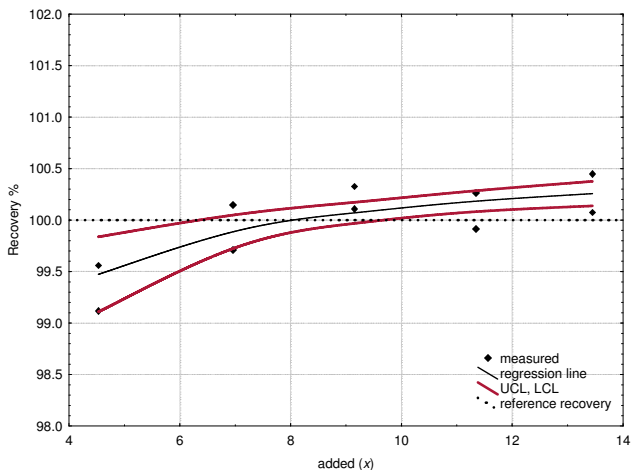


Figure 2. Confidence interval for the recovery

point calibration (if it is less than allowed). This can be evaluated by calculating a confidence interval for the bias from the parameters of the regression line fitted on method validation data. The mean confidence interval is shown in Figure 3 calculated for 5000 generated lines. This reflects the average properties of the intervals. It can be seen that the bias depends on the unknown concentration (x) with fixed standard concentration (x^* , set during single-point calibration). If the standard equals to the unknown concentration, the single-point calibration is genuinely unbiased ($x = 100$).

Figure 3 shows cases with non-zero true intercept (β_0), causing the true bias (the center of the intervals) to be dependent on the difference of the standard and unknown concentrations. If the true intercept is zero, then the true bias is zero independently of the concentrations, and the center of the intervals is horizontal. It can be seen that increasing variance causes wider intervals (solid and dashed lines), meaning lower probability of acceptance of single-point calibration.

To compare the procedures for the content uniformity (IV) assessment, Table 1 shows the tolerance interval widths of the current criteria of the Ph. Eur. and the proposed calculation method. The first column shows the measured CU relative standard deviations (RSDs), the second one shows the interval widths with the Ph. Eur. method, while the other columns show the interval widths obtained with the correct calculation using the analytical error (RSD) in the header. The 3rd column reflects the case when the sources of variation are not separated, but the correct tolerance factor is used for the interval calculation.

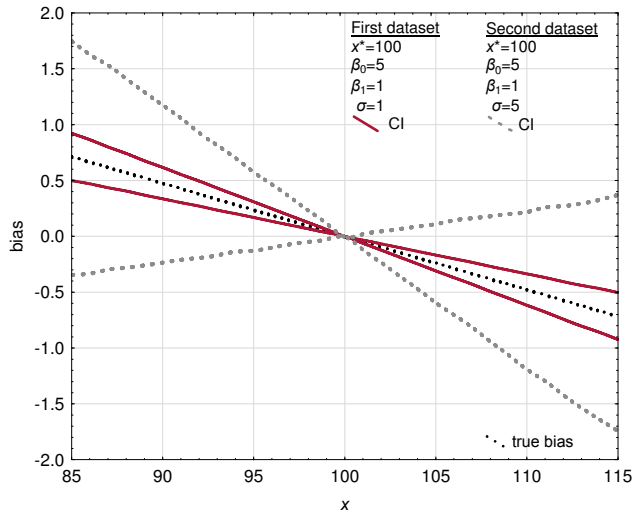


Figure 3. Simulated confidence bands for the single-point calibration

Figures typeset in bold are not in compliance with the criterion of $\pm 15\%$ of the current regulation.

It is clear that the pharmacopeial method is too liberal; the content uniformity is accepted with all the included standard deviation values. When accepting the uniformity, we may not confidently state that the majority of the tablets have an active content in the $\pm 15\%$ range of the label claim, because of the too small tolerance factor. In contrast, it can be stated about my proposed intervals that at least 99% of the product batch have an active content deviating from the label claim no more than $\pm 15\%$, with 95% probability (if they are contained in the $\pm 15\%$ range).

It can be seen that some cases exist where overlooking the two sources of variation causes wrong conclusion (e. g. 3.5 or 4% content uniformity RSD).

Table 1. One-sided width of the tolerance intervals (ks)

Measured RSD%	Ph. Eur. $k = 2.40$	One-way random model: analytical method RSD%					
		0.25	0.50	1.00	1.50	2.00	3.00
0.50	1.20	2.22					
1.00	2.40	4.44	3.63				
1.50	3.60	6.67	6.15	4.32			
2.00	4.80	8.89	8.51	7.27	4.64		
3.00	7.20	13.33	13.08	12.30	10.90	8.64	
3.50	8.40	15.56	15.34	14.68	13.52	11.74	3.94
4.00	9.60	17.78	17.59	17.02	16.02	14.54	9.28
5.00	12.00	22.22	22.07	21.62	20.84	19.71	16.13

5 NEW SCIENTIFIC RESULTS

1. In the topic of analytical method transfer: [P1, P5, P8]
 - (a) I have compared the traditional hypothesis of no difference and the two one-sided t -tests (TOST) and concluded that the two methods are equal, provided the sample size is adequate. In this case both the type I and type II errors are maintained on the same level with both hypotheses. In accordance with the literature, I have shown, when the sample size is lower than required, the type II error of the hypothesis of no difference (the consumer's risk) can become large. The TOST provides protection against this kind of error: it does not permit the acceptance, if it is not justified.
 - (b) I have discussed the risks when the analytical variance is not known — which is the usual case in the pharmaceutical analysis — and it is underestimated. The traditional hypothesis controls the manufacturer's risk, while the TOST maintains a fixed consumer's risk.
 - (c) I have provided a method for calculating the number of tablets to be homogenized to make practically identical samples in the two laboratories.
2. In the topic of accuracy of analytical methods: [A2, P13]
 - (a) I have shown that the usual regression method to determine the accuracy of analytical methods is flawed. It investigates the linear relationship between the added and recovered (measured) concentration by checking the zero intercept and unit slope, but the non-independence of these two estimated parameters is overlooked.
 - (b) I have demonstrated that even the simultaneous assessment of the intercept and slope does not answer the relevant question. Accepting a zero intercept and unit slope does not prove that the analytical method is accurate, just the existence of bias is failed to be justified. The consumer's risk is not fixed, only controlled by the number of observations.
 - (c) I have developed a method based on the two one-sided hypotheses for the evaluation of the accuracy. This new method answers the relevant question, if it is considered to be proven that the bias does not exceed an allowed value. The consumer's risk is controlled at all times.

3. In the topic of assessing the validity of single-point calibration: [A4, P9, P11]
 - (a) I have shown that the usual practice of assessing the applicability of this simplification is flawed. Testing the zero intercept of the linear relationship between the analytical signal and the concentration is not statistically sound, because failing to reject a zero intercept does not mean it is in fact zero. Moreover, testing an interval hypothesis for the intercept is correct in a statistical sense, but not relevant, as the user is particularly interested in the bias of the calculated concentration (x scale) and not in the analytical signal (y scale).
 - (b) I have developed a method for assessing the relevant question. With my procedure the user is able to demonstrate that the bias in the calculated concentration does not exceed a predefined limit (stated in concentration units).

4. In the topic of content uniformity: [A1, A3, P2, P4, P7]
 - (a) By examining the regulations (pharmacopeias and regulatory guidelines), I have shown and demonstrated by computations that the current criteria are flawed. They are based on tolerance intervals which are relevant for the practical question, however, the calculation method is incorrect: it does not care for the two sources of variability in the content uniformity measurements, namely the product inhomogeneity and the analytical error. The consequence is that the interval which is supposed to contain a proportion (e. g. 90%), is in fact narrower: the interval contains only a smaller proportion.
 - (b) I have proposed a calculation method based on the tolerance interval for a variance component model. This method is capable of separating the analytical error from the product inhomogeneity, and constructs the tolerance interval solely for the inhomogeneity.
 - (c) The assay and the content uniformity — as two different criteria that have to be assessed concurrently — are not consistent, which has been shown and demonstrated by calculations.
 - (d) I have proposed a holistic approach for treating the inhomogeneity and the mean active content simultaneously, exploiting the fact that the mean and variance are statistically independent. As an alternative I have discussed the applicability of Taguchi's loss function as well.

6 APPLICATIONS

The cases mentioned in my dissertation come from the pharmaceutical practice, where the interpretation of the results of traditional hypothesis tests cause serious issues.

The proposed procedure for the assessment of method transfer (I) proved to be effective, both from the perspective of the hypothesis test, and from the preparation of identical samples. Some departments of the Gedeon Richter Pharmaceutical Works use these procedures.

The proposal for the assessment of the accuracy (II) of analytical methods is easily applicable as well, the confidence interval for the fitted line can be computed with the used statistical packages used in routine analysis. The proposed method gives correct and convincing results; the accuracy is justified, if it is accepted.

Similarly, the method developed for the single-point calibration (III) gives appropriate answer for the relevant question: the decision is based on the bias caused by applying the single-point calibration. The formulas of the calculation use the parameters of the fitted line measured during validation, these can be included in e. g. Excel tables. These two procedures are under investigation at a department of Gedeon Richter Plc.

The content uniformity (IV) evaluation is a current Pharmacopeia procedure, so a quick change is less likely in that case. The statistically sound method gives useful results, but requires advanced computations, which cannot be performed with office software. Applying tabulated tolerance factors for different inhomogeneity and analytical error could be a solution for this problem.

7 PUBLICATIONS

JOURNAL ARTICLES OF THE DISSERTATION

- [A1] Bánfai B., Ganzler K. and Kemény S. (2007) Content uniformity and assay requirements in current regulations. *Journal of Chromatography A* 1156: 206–212. IF: 3.641, CIT: 2 doi:10.1016/j.chroma.2006.10.067
- [A2] Kemény S., Deák A. and Bánfai B. (2009) Testing accuracy of analytical methods by regression. *Journal of Chemometrics* 23: 211–216. IF: 1.291, CIT: 2 doi:10.1002/cem.1219
- [A3] Komka K., Kemény S. and Bánfai B. (2010) Novel tolerance interval model for the estimation of the shelf life of pharmaceutical products. *Journal of Chemometrics* 24: 131–139. IF: 1.377 doi:10.1002/cem.1294

- [A4] Bánfai B. and Kemény S. (2012) Estimation of bias for single-point calibration. *Journal of Chemometrics* In Press. IF: 1.377 doi:10.1002/cem.2417

OTHER JOURNAL ARTICLES

- [O1] Bánfai B., Jia H., Khatun J., Wood E., Risk B., Gundling W., Kundaje A., Gunawardena H.P., Yu Y., Xie L., Krajewski K., Strahl B.D., Chen X., Bickel P.J., Giddings M.C., Brown J.B. and Lipovich L. (2012) Long non-coding RNAs are rarely translated. *Genome Research* Under Review. IF: 13.588
- [O2] The ENCODE Project Consortium (2012) An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* Under Review. IF: 36.101

PRESENTATIONS AND POSTERS

- [P1] Bánfai B., Deák A. and Kemény S. (2006) *Statistical Background of Analytical Method Transfer* in: 3rd Symposium on Computer Applications and Chemometrics in Analytical Chemistry, Tihany, Hungary. PRESENTATION IN ENGLISH
- [P2] Bánfai B., Ganzler K. and Kemény S. (2007) *Statistically Sound Proposals for Content Uniformity Test of Solid Drug Products* in: Dalian International Symposia and Exhibition on Chromatography (DISEC 2007), Dalian, China. PRESENTATION IN ENGLISH
- [P3] Bánfai B., Ganzler K. and Kemény S. (2009) *Statistical Aspects of Intermediate Precision Studies* in: Spring Research Conference on Statistics in Industry and Technology, Vancouver, Canada. PRESENTATION IN ENGLISH
- [P4] Bánfai B. and Kemény S. (2006) *Toleranciaintervallumok számítása a gyógyszeripari minőségbiztosítás során* in: Alkalmazott Informatikai Konferencia, Kaposvár, Hungary. PRESENTATION IN HUNGARIAN
- [P5] Bánfai B., Deák A. and Kemény S. (2006) *Az analitikai módszerátadás statisztikai háttere* in: 34. Műszaki Kémiai Napok, Veszprém, Hungary. PRESENTATION IN HUNGARIAN
- [P6] Bánfai B., Ganzler K. and Kemény S. (2007) *Gyógyszerkészítmények átlagos és egyedi hatóanyag-tartalmának statisztikai vizsgálata* in: IV. Doktoráns Konferencia, Oláh György Doktori Iskola, BME Vegyész-mérnöki és Biomérnöki Kar, Budapest, Hungary. PRESENTATION IN HUNGARIAN
- [P7] Bánfai B., Ganzler K. and Kemény S. (2006) *Content Uniformity and Assay Requirements in Current Regulations – From the Industry's Perspective* in: The 30th International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC 2006), San Francisco, CA, USA. POSTER
- [P8] Bánfai B., Deák A., Ganzler K. and Kemény S. (2007) *Statistical Background of Analytical Method Transfer* in: Pharmaceutical Sciences World Congress, Amsterdam, The Netherlands. POSTER
- [P9] Bánfai B. and Kemény S. (2010) *Evaluation of Single-point Calibration During Analytical Method Validation* in: The 35th International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC 2010), Boston, MA, USA. POSTER

- [P10] Kemény S., Deák A. and Bánfai B. (2007) *Paradigmatic change in decision based on chemical analysis* in: Conferentia Chemometrica, Budapest, Hungary. PRESENTATION IN ENGLISH
- [P11] Kemény S. and Bánfai B. (2009) *Estimation of bias for single-point calibration* in: Conferentia Chemometrica, Siófok, Hungary. PRESENTATION IN ENGLISH
- [P12] Kemény S., Deák A. and Bánfai B. (2007) *A kémiai analízisen alapuló döntéshozatal új paradigmája* in: MKE Centenárium Konferencia, Sopron, Hungary. PRESENTATION IN HUNGARIAN
- [P13] Kemény S., Deák A. and Bánfai B. (2009) *Az analitikai módszerek torzítatlanságának megítélése regressziós vizsgálattal* in: KeMoMo-QSAR, Szeged, Hungary. PRESENTATION IN HUNGARIAN
- [P14] Kemény S., Deák A. and Bánfai B. (2010) *Paradigmaváltás a hipotézisvizsgálatnál* in: KeMoMo-QSAR, Szeged, Hungary. PRESENTATION IN HUNGARIAN
- [P15] Jia H., Bánfai B., Khatun J., Maier C. W., Bickel P. J., Giddings M. C., Lipovich L., Brown J. B., ENCODE AWG and ENCODE Consortium (2011) *Mass spectrometric analysis demonstrates that few human lncRNAs are translated* in: Genome Informatics, Cold Spring Harbor, NY, USA. POSTER