

Accepted Manuscript

Robustness testing in the determination of seven drugs in animal muscle by liquid chromatography–tandem mass spectrometry

M.L. Oca, L. Rubio, M.C. Ortiz, L.A. Sarabia, I. García

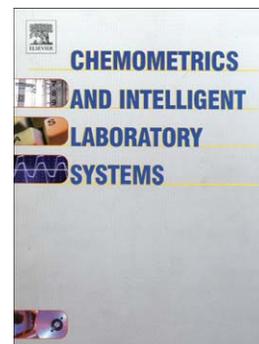
PII: S0169-7439(15)00334-2
DOI: doi: [10.1016/j.chemolab.2015.12.019](https://doi.org/10.1016/j.chemolab.2015.12.019)
Reference: CHEMOM 3158

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received date: 16 March 2015
Revised date: 26 November 2015
Accepted date: 29 December 2015

Please cite this article as: M.L. Oca, L. Rubio, M.C. Ortiz, L.A. Sarabia, I. García, Robustness testing in the determination of seven drugs in animal muscle by liquid chromatography–tandem mass spectrometry, *Chemometrics and Intelligent Laboratory Systems* (2016), doi: [10.1016/j.chemolab.2015.12.019](https://doi.org/10.1016/j.chemolab.2015.12.019)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



ROBUSTNESS TESTING IN THE DETERMINATION OF SEVEN DRUGS IN ANIMAL MUSCLE BY LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY

M.L. Oca^a, L. Rubio^a, M.C. Ortiz^{a,1}, L.A. Sarabia^b, I. García^c

^a *Department of Chemistry*, ^b *Department of Mathematics and Computation*
Faculty of Sciences, University of Burgos
Pza. Misael Bañuelos s/n (09001) Burgos, Spain.

^c *Laboratorio de Salud Pública de Burgos*
S.T. de Sanidad y Bienestar Social (Junta de Castilla y León)
Paseo Sierra de Atapuerca, 4 (09071) Burgos, Spain

Abstract

In this work, the robustness of the sample preparation procedure for the determination of six tranquillizers (xylazine, azaperone, propionylpromazine, chlorpromazine, haloperidol and azaperol) and a beta-blocker (carazolol) in animal muscle by LC/MS-MS was assessed through the experimental design methodology. A 2_{III}^{7-4} fractional factorial design was performed to evaluate the influence of seven variables on the final concentration of the seven drugs in the samples, in accordance with what is laid down in Commission Decision No 2002/657/EC. The variation considered for each of those seven factors is likely to happen when preparing the samples, being the values chosen as level -1 the nominal operating conditions. The results of the experimentation were evaluated from different statistical strategies, such as hypothesis testing using an external variance previously estimated, Lenth's method and Bayesian analysis. Both Lenth's and Bayes' approaches enabled to determine the effect of every variable even though no degrees of freedom were left to estimate the residual error. The same conclusion about the robustness of the extraction step was reached from the three methodologies, namely, none of the seven factors examined influenced on the method performance significantly, so the sample preparation procedure was considered to be robust.

Keywords

Robustness; tranquillizer; liquid chromatography-tandem mass spectrometry; Commission Decision No 2002/657/EC; Lenth's method; Bayesian analysis.

Abbreviations²**1. Introduction**

Assessing the potential sources of variability in one or several responses of an analytical procedure must be a key part of method development. This involves making

¹ Corresponding author. Telephone number: +34 947259571; fax: +34 947258831. *E-mail address:* mcortiz@ubu.es (M.C. Ortiz).

² International conference on harmonization of technical requirements for Registration of pharmaceuticals for human use (ICH), International Union of Pure and Applied Chemistry (IUPAC), liquid chromatography coupled to tandem mass spectrometry (LC/MS-MS), xylazine (XYL), azaperone (AZA), propionylpromazine (PROP), chlorpromazine (CHLOR), haloperidol (HAL), azaperol (AZOL), carazolol (CAR), solid-phase extraction (SPE), maximum residue limit (MRL), internal standard (IS), atmospheric pressure chemical ionization (APCI), positive ion electrospray (ESI⁺), multiple reaction monitoring (MRM), least squares (LS), variance inflation factor (VIF).

deliberate and small changes in nominal experimental conditions and investigating their subsequent effect on performance to identify the variables with the most significant effect and ensure that they are closely controlled when using the method [1].

From this perspective, two terms referring to the evaluation of the method performance still coexist within the scientific vocabulary: robustness and ruggedness. They have often been used as synonyms [2,3,4,5], but a distinction between both has also been drawn in accordance with their information about different features of an analytical method: its practicability and stability related to experimental physicochemical variables that are internal to the method (robustness) and its interlaboratory transferability when the variables under study are external to the method (ruggedness) [6,7,8]. In this sense, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) defines the robustness of an analytical procedure as "a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage" [9]. On the other hand, the United States Pharmacopeia and The National Formulary has adopted the ICH definition of robustness and defines the ruggedness of an analytical method as "the degree of reproducibility of test results obtained by the analysis of the same samples under a variety of conditions such as different laboratories, analysts, instruments, lots of reagents, elapsed assay times, assay temperatures, or days " [10]. However, there is still some confusion in scientific journals, guidelines and monograph literature regarding the use of these words when applied to analytical methods [11].

Information about ruggedness/robustness should be indicated in the laboratory procedure [1]. Anyway, the strategy to be followed in a robustness and/or ruggedness test is the same. It involves performing a screening study usually by means of experimental designs after the identification of the potentially influential factors and the definition of their variation ranges and of the responses to be determined. At this point, conducting either a Plackett–Burman design [12] or a fractional factorial design, as in the Youden test [13], is the most frequently used procedure for robustness/ruggedness evaluation. The choice of the design to be performed depends on the purpose of the test and on the number of factors to be examined [4]. Due to the minimum time and analytical effort required, Commission Decision (EC) No 2002/657 [14] encourages the application of the Youden approach to the compulsory verification of that performance characteristic both in screening and confirmatory methods for the monitoring of certain substances and residues thereof in live animals and animal products. Many examples in this field can be found in the literature [6,15,16,17,18,19]. The Eurachem Guide to Method Validation and Related Topics [1] recommends, whenever possible, the evaluation of the ruggedness/robustness of a method by using the Youden test. IUPAC [2] also recognizes the strategy described by Youden as adequate to study the ruggedness of an analytical method.

This work shows the evaluation of the performance of the sample preparation step prior to the simultaneous determination of seven drugs in animal muscle by liquid chromatography coupled to tandem mass spectrometry (LC/MS-MS). More precisely, the substances analysed were five tranquillizers (xylazine (XYL), azaperone (AZA), propionylpromazine (PROP), chlorpromazine (CHLOR) and haloperidol (HAL)), one of the metabolites of azaperone (azaperol (AZOL), which is derived from the former by reduction) and a blocker agent of the β -adrenergic receptor (carazolol (CAR)). Their chemical structures are depicted in Fig. 1. As the sample preparation stage includes sampling, pretreatment and solid-phase extraction (SPE) steps, it will be quite likely to be responsible of the highest errors in the determination. So, the effect of seven factors related to the sample preparation procedure on the final concentration of every drug in the sample was examined through an eight-experiment Youden design. These factors

were deliberately changed between nominal and extreme conditions that represented the variability that may well occur when performing routine analyses. The results arising from the experimental plan were interpreted from several statistical methodologies in order to assess the robustness of the extraction step.

As the proposed design was saturated, an independent estimation of the experimental error as standard deviation at a previous stage of the method development was used to evaluate the significance of the factors. In addition, Lenth's and Bayes' approaches [20] have also been applied for drawing conclusions on the robustness/ruggedness of the sample preparation stage.

Veterinary medicinal products are necessary to ensure animal health and welfare, but their administration to food-producing animals may leave residues in them. This is the case of sedatives, which are often used in animal production, especially in pigs. These are more aggressive than other farm animals and particularly sensitive to stress during handling and, more specifically, during their transport to the slaughterhouse. Not palliating all this stress will result in high premature mortality and meat of poorer quality called Pale Soft Exudative [21,22], so pigs are usually injected with sedatives a few hours before slaughtering to calm them down. Due to their potential effect on the activity of the human nervous system, residues of these drugs in foodstuffs of animal origin constitute a hazard to human health, which makes their analytical control necessary. Consequently, European legislation on this matter has been developed. As part of it, Commission Regulation (EU) No 37/2010 [23] establishes maximum residue limits (MRLs) for some allowed pharmacologically active substances depending on the animal species and on the target tissue according to the procedures in Regulation (EC) No 470/2009 [24]. Prohibited pharmacologically active substances for which a MRL cannot be established are also listed in [23]. In the specific case of the sedatives analysed in this study, the MRL for azaperone, as the sum of azaperone and azaperol, is $100 \mu\text{g kg}^{-1}$, and that for carazolol is $5 \mu\text{g kg}^{-1}$, both values in porcine muscle tissue, while the use of chlorpromazine in food-producing animals is not allowed.

The rules laid down in Commission Decision No 2002/657/EC [14] require confirmatory methods for organic residues such as those from veterinary drugs or contaminants to provide information on the chemical structure of the analyte. In fact, the European Union considers mass spectrometry the most suitable analytical tool for a correct identification of these compounds [25]. When full-mass spectra are not recorded, a system of identification points with different requirements for prohibited substances and for those with MRLs (listed in Group A or B, respectively, of Annex I of Directive 96/23/EC) shall be used to interpret the data.

2. Theory

Full factorial designs are useful for experimenting with relatively few factors. As the number of factors in a 2^k factorial design increases, the number of runs required for its consecution rapidly outgrows the resources of most experimenters. But it is possible to run only a fraction of the full factorial set that provides nearly as much information as in the corresponding full factorial design if it can be reasonably assumed that certain high-order interactions are negligible. These fractional factorial designs are among the most widely used types of designs for product and process design and improvement, and industrial/business experimentation [26], being also greatly applied in screening studies such as those for evaluating the robustness/ruggedness of an analytical method.

A 2^k fractional factorial design containing 2^{k-p} runs is called a $1/2^p$ fraction of the 2^k design or, more simply, a 2^{k-p} fractional factorial design, where k is the number of factors considered and p the number of independent generators selected for the

construction of the design, being $p < k$. More information about the construction of fractional factorial designs can be found in [26,27].

The resolution of a fractional factorial design shows how the estimated effects are confounded. In particular, resolution III fractional factorial designs (2_{III}^{k-p}) are those in which no main effects are aliased with any other main effect, but they are aliased with two-factor interactions and some two-factor interactions may be aliased with each other [26]. So, as main effects can be estimated only in the absence of interaction effects, these will be considered to be negligible when estimating those.

Two-level resolution III designs can be constructed to evaluate $k = (N - 1)$ factors with N runs, being N a power of 2. An example of resolution III fractional factorial designs is the robustness test introduced by Youden and Steiner for analytical chemistry [13]; this kind of experimental designs was well-established in other fields of science, though, such as in agriculture since the 1920s [28].

The Youden robustness test makes use of a two-level 2_{III}^{7-4} fractional factorial design to study the influence of up to seven factors in eight experiments. This is the strategy recommended by different international regulatory bodies to check the robustness/ruggedness of an analytical method [1,2,13,14,29]. For the assignment of the two levels of every factor, although a symmetrical interval around every nominal operating condition (encoded as 0) between two extreme levels (encoded as -1 and $+1$, respectively) is often set, it is also stated in [1,2,13,14,29] that the different influential variables should be changed from their nominal values (encoded as -1) to those extreme (encoded as $+1$) meaning a slight change with regard to the nominal operating conditions. The existence of curvature in any of the monitored factors is not expected because the variation from level -1 to level $+1$ is quite small.

When seven factors are evaluated from eight runs performed following the experimental plan of the 2_{III}^{7-4} fractional factorial design proposed by Youden and Steiner, as no interactions are taken into account, the mathematical model posed to express the underlying relation between the response under study y and the seven factors considered is, in codified variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \varepsilon \quad (1)$$

where ε is the experimental error and the variable x_i , $i = 1, \dots, 7$, equals either -1 or $+1$ whether the i -th factor is at either nominal or extreme level, respectively.

Since the model in Eq. (1) consists of eight coefficients and only eight runs are available, the effect of every factor will thus be calculated rather than estimated as follows:

$$\beta_i = \frac{\sum y(+1) - \sum y(-1)}{4} \quad (2)$$

where $\sum y(+1)$ and $\sum y(-1)$ represent the sums of the responses where factor x_i is at $(+1)$ and (-1) level, respectively. There will be no degrees of freedom left to estimate the experimental error ε either, so the evaluation of the statistical significance of either the model in Eq. (1) or its coefficients by a hypothesis test will not be possible from the design. At this point, several alternatives are available.

2.1. Using an external variance for the estimation of the residual error

The first choice is to have an estimation of the experimental error as standard deviation s , which could have been calculated either from previous tests with the analytical procedure of interest or by performing more experiments (replicates of one or more design points). Both choices will make it possible to estimate the significance of the coefficients in Eq. (1) thanks to the following hypothesis test:

$$H_0: \beta_i = 0 \text{ (There is not effect of the factor } i) \quad (3)$$

$$H_a: \beta_i \neq 0 \text{ (There is effect of the factor } i)$$

The critical region (CR) of this hypothesis test, at a 95% confidence level, is defined as

$$CR = \left\{ \frac{\beta_i}{s_{\beta_i}} < -1.96 \text{ or } \frac{\beta_i}{s_{\beta_i}} > 1.96 \right\} \text{ for } i = 1, \dots, 7 \quad (4)$$

where the value of the standard deviation s_{β_i} is calculated from Eq. (5), being n the number of runs in the experimental design ($n = 8$ in this work) and s the experimental error externally estimated.

$$s_{\beta_i} = \frac{s}{\sqrt{n}} \quad (5)$$

2.2. Lenth's method

Another alternative means the use of statistical approaches that do not need a previous estimation of the residual error. The standard procedure for the analysis of an unreplicated two-level factorial design is the normal (or half-normal) plot of the estimated factor effects. However, unreplicated designs are so widely used in practice that many formal analysis procedures have been proposed to overcome the subjectivity of the normal probability plot [26]. One of these options is Lenth's method [20,26]. This is based on the idea that, if none of the factors is active, the coefficients of the model in Eq. (1), except for the constant term β_0 , will have values around 0, whereas the coefficient of an active factor will be different from 0. So, if the absolute values of those coefficients are considered, an initial estimate of the coefficient standard deviation s_{β} is obtained by arranging those absolute values in increasing order and multiplying their median by 1.5. Any coefficient whose absolute value is greater than $2.5s_{\beta}$ is then removed from the list and the whole procedure is repeated until no further coefficient is rejected. Let c denote the number of coefficients that remain in the end. A coefficient will be active at a 5% significance level if it lies out of the interval $(-t_{0.025, c/3} s_{\beta}, t_{0.025, c/3} s_{\beta})$, being $t_{0.025, c/3}$ the value of a Student's t -distribution for a probability equal to 0.025 and $c/3$ degrees of freedom.

2.3. Bayesian approach

In line with Section 2.2, a second option when no estimation of the experimental variance is possible is the Bayesian analysis of the coefficients β_i [20]. This procedure is based on calculating the *a posteriori* probability that the respective factor is active. This involves that: i) only a fraction α (between 0.10 and 0.40) of the factors is active; ii) inactive coefficients are assumed to follow a normal distribution with mean 0 and constant variance; and iii) the coefficients of the active factors are also supposed to be normally distributed, but now with a higher variance, so that the ratio k of the variance of the active factors to that of the inactive ones ranges from 5.0 to 15.0. By changing α and k , it is possible to determine both the maximum and the minimum *a posteriori* probabilities of every factor being active.

3. Material and methods

3.1. Reagents and chemicals

HPLC-grade methanol, analytical-grade ethanol, HPLC-grade acetonitrile, formic acid (98-100% purity), 32% ammonia solution and analytical-grade potassium dihydrogen phosphate were all purchased from Merck (Darmstadt, Germany). Analytical-grade orthophosphoric acid and potassium hydroxide (both 85% purity) were supplied by Panreac (Barcelona, Spain).

A solution of phosphate buffer (0.1 M, pH 6, 1 L) was prepared by dissolving potassium dihydrogen phosphate into Milli-Q water and adjusting pH to 6 with 10 M potassium hydroxide.

Azaperol (97% minimum purity) was supplied by Dr. Ehrenstorfer GmbH (Augsburg, Germany). Xylazine hydrochloride, haloperidol, chlorpromazine hydrochloride, propionylpromazine hydrochloride, azaperone and azaperone-d₄ (used as the internal standard (IS) for both azaperone and azaperol) were obtained from Sigma-Aldrich (Steinheim, Germany), all of them with a 98% minimum purity. Chlorpromazine-d₃ (IS for both chlorpromazine and propionylpromazine) and haloperidol-d₄ (IS for haloperidol, carazolol and xylazine) were purchased from LGC Standards (Teddington, UK), both with a 98% minimum purity. Carazolol (Suacron®, 98% minimum purity) was a generous gift from Divasa-Farmavic, S.A. (Barcelona, Spain). Whenever needed, ultra-pure water was obtained from a Milli-Q water purification system (Millipore, Bedford, MA, USA).

3.2. Standard solutions

Stock solutions of xylazine, azaperone, propionylpromazine, chlorpromazine, haloperidol, azaperol and carazolol at 100 mg L⁻¹ and individual stock solutions of azaperone-d₄ (100 mg L⁻¹), chlorpromazine-d₃ (1 mg L⁻¹) and haloperidol-d₄ (1 mg L⁻¹) were prepared in ethanol and stored in amber bottles at 4 °C. Concentrations of these analytes must be referred to their free forms, not to the hydrochloride ones.

Two working solutions were prepared by further dilution in ethanol: one containing the seven non-deuterated sedatives at the following concentrations: 100 µg L⁻¹ for xylazine, 2500 µg L⁻¹ for azaperone, 150 µg L⁻¹ for propionylpromazine, 100 µg L⁻¹ for chlorpromazine, 50 µg L⁻¹ for haloperidol, 2000 µg L⁻¹ for azaperol and 125 µg L⁻¹ for carazolol; and the other with the three ISs at concentrations of 2500 µg L⁻¹ for azaperone-d₄, 200 µg L⁻¹ for chlorpromazine-d₃ and 40 µg L⁻¹ for haloperidol-d₄. Both solutions were stored in amber bottles at 4 °C.

3.3. Sample pretreatment and purification procedure

For the evaluation of the seven influential factors listed in Table 1, eight muscle samples were prepared according to the experimental plan of the Youden design shown in Table 2 (in codified variables). In addition, six matrix-matched calibration standards including a blank underwent the same pretreatment process, being the 7 variables selected as factors set at their nominal conditions (level -1). In all cases, 5 g of homogenised animal muscle was weighed in a 50 mL polypropylene tube and then fortified with the three ISs by adding a 50 µL aliquot of the working internal standard solution. Next, 100 µL of the working standard solution of the sedatives was added to all robustness samples, while five of the six matrix-matched standards were spiked, respectively, with 50, 100, 200, 300 and 400 µL of that solution. Before the preparation of both the samples and the standards, it had been checked that the animal muscle tissue used as matrix was completely free of the analytes under study. This avoided performing additional experimentation for the calculation of recovery.

10 mL (level -1) or 9 mL (level +1) of a newly prepared (level -1) or a one-month-aged (level +1) phosphate buffer solution was pipetted to every sample and the mixture was stirred in a vortex mixer for 30 s. After 10 min in an ultrasonic bath, 3 mL of 1% orthophosphoric acid was added, and every sample was vortex mixed again for 30 s and then centrifuged for 10 min at 6800 rpm at room temperature. The supernatant of every homogenate was clarified using a Millex-AP prefilter coupled to a 20 mL syringe.

The purification step by solid-phase extraction (SPE) was performed either manually (level -1) or automatedly (level +1). SPE cartridges were first conditioned with 3 mL of methanol and then equilibrated with 3 mL of Milli-Q water. After every sample filtrate had been loaded, cartridges were rinsed with 3 mL of 1 M formic acid and next with 3 mL (level -1) or 3.5 mL (level +1) of a methanol/1 M formic acid (55/45 v/v) washing solution. After drying by passing an air stream under vacuum for 10 min, elution of the analytes and the internal standards was performed with 4 mL (level -1) or 4.5 mL (level +1) of a newly prepared (level -1) or one-week-aged (level +1) eluting solution (ammonium hydroxide/ethanol 2/98 v/v). Each eluate was collected in a 15 mL amber tube, evaporated to dryness under nitrogen at 38 °C, reconstituted with 250 µL of 30% acetonitrile and placed in an ultrasonic bath for 5 min. Every final extract was clarified through a 0.22-µm nylon filter coupled to a 1 mL syringe, and the filtrate was transferred to an insert contained into a vial for LC/MS-MS analysis.

This sample pretreatment and purification procedure had been optimized prior to carrying out the robustness study described in this work. The thorough cleanup step performed before LC/MS-MS analysis, together with the use of three deuterated internal standards and matrix-matched calibration, helped to avoid the ionization-suppression/enhancement problem as far as possible.

3.4. Instrumental

Millex-AP prefilters were obtained from Millipore Corporation (Bedford MA, USA). A Jouan C3i centrifuge (Thermo Electron Corporation, San Jose, CA, USA) was used. The vacuum manifold used for the SPE step was purchased from Waters Corporation (Milford, MA, USA). SPE cartridges were 300 mg/6 mL Varian Bond Elut LRC Certify® (Agilent Technologies, Waldbronn, Germany).

3.5. LC/MS-MS analysis

Chromatographic separation was carried out on an Agilent 1200RR series liquid chromatograph (Agilent Technologies, Waldbronn, Germany) equipped with a binary pump, a degasser, an autosampler and a column heater, in which a Gemini NX C18 column (50 mm×2mm I.D.; particle size, 3 µm; Phenomenex, Macclesfield, UK) was placed at 30 °C. Two mobile phases were used in a gradient program, in which the mobile phase A consisted of acetonitrile containing 0.5% ammonium hydroxide and mobile phase B consisted of water containing 0.5% ammonium hydroxide. The initial composition was 30% A and 70% B, which was increased linearly to 85% A in the first 7 min, held at these conditions for 2 min and returned to the initial composition from minute 9 to minute 11. The column was then equilibrated in these conditions for 4 min before the next injection, being the total run time 15 min. The flow rate was set at 0.3 mL min⁻¹ and the injection volume was 10 µL.

Tandem mass spectrometry was performed on an Agilent 6410B triple quadrupole mass analyzer with a multimode ionization source which was programmed to operate in mixed conditions of atmospheric pressure chemical ionization (APCI) and positive ion electrospray (ESI⁺). The source temperature was held at 250 °C and the capillary voltage was set at 4300 V. Nitrogen was used both as nebuliser and collision gas. Analytes were detected using the multiple reaction monitoring (MRM) mode with a

dwelt time of 70 ms per ion. Both Q1 and Q3 quadrupoles were maintained at unit resolution. As can be seen in Table 3, for the confirmation of every drug, its molecular ion was selected as the precursor ion and two transitions thereof were monitored, while quantification was carried out with the most intense transition. Regarding the three internal standards, only one transition from its corresponding molecular ion was considered in each case.

3.6. Software

The experimental design performed was constructed and interpreted using NEMRODW [30]. All the regression lines were estimated and validated using STATGRAPHICS Centurion XVI [31].

4. Results and Discussion

As one of the last stages in the development of a method for the multiresidue determination of seven drugs in animal muscle by LC/MS-MS, the robustness of the sample preparation step was assessed by means of the strategy proposed by [13] and recommended by the legislation in force in the European Union [14]. The seven experimental factors evaluated were: 1) Analyst; 2) Way of performing the SPE stage by using either a multistation vacuum manifold or an automated extraction system; 3) Time elapsed between the preparation of the eluting solution, whose expiry was set to 1 week, and the day of analysis; 4) Time elapsed between the preparation of the phosphate buffer solution, whose expiry was set to 1 month, and the day of analysis; 5) Volume of the phosphate buffer added to the samples; 6) Volume of the ammonia/ethanol eluent in SPE; 7) Volume of the methanol/formic acid rinsing solution in SPE. Table 1 shows these seven factors as well as the two levels chosen for each of them. As commented in Section 2 and in accordance with the procedures in [1,2,13,14,29], the nominal operating conditions of the method were selected as level -1 for every factor.

Seven experimental responses were considered, namely the concentration of the seven drugs in a muscle sample that had been fortified with 100 μL of the working standard solution. The concentration of each sedative in every muscle sample was determined from the corresponding least squares (LS) linear regression of standardized peak area (calculated as the ratio between the absolute peak area of the analyte of interest and that of its internal standard) versus true concentration. These seven regression models were estimated from a calibration set of six matrix-matched standards, each of them corresponding, besides a blank, to the addition of 50, 100, 200, 300 and 400 μL of the working standard solution, respectively. The calibration range of every analyte was thus 0 – 6.92 $\mu\text{g kg}^{-1}$ for xylazine, 0 – 202.00 $\mu\text{g kg}^{-1}$ for azaperone, 0 – 11.15 $\mu\text{g kg}^{-1}$ for propionylpromazine, 0 – 7.26 $\mu\text{g kg}^{-1}$ for chlorpromazine, 0 – 4.00 $\mu\text{g kg}^{-1}$ for haloperidol, 0 – 160.00 $\mu\text{g kg}^{-1}$ for azaperol, 0 – 10.03 $\mu\text{g kg}^{-1}$ for carazolol. The matrix-matched standards were pretreated and analysed under the nominal operating conditions of the method (see Sections 3.3 and 3.5).

The influence of the seven factors considered on the final concentration of every drug in a muscle sample was studied through a 2_{III}^{7-4} fractional factorial design consisting of eight experiments. As can be seen in Table 2, the resulting experimental matrix matches up with that of the Youden proposal by multiplying by (-1) every element of the matrix. The suitability and great quality of this experimental design to be performed is well-known. As its information matrix (and thus its dispersion matrix) is a diagonal matrix, the orthogonality of the design is ensured. Because of this orthogonal design structure, all main effects would be estimated independently, and the variance inflation

factors (VIFs) of all the coefficients in the model of Eq. (1), which describes the relationship between every response and the seven factors monitored, were equal to 1. As this is the best value for a VIF, the estimates of those coefficients would be the most precise possible ones.

Once the whole experimentation had been carried out, the presence of both all the analytes and the internal standards in every sample was confirmed from their precursor-to-product transitions, listed in Table 3. Then, the standardized peak area of every analyte in every run was calculated from its most intense transition, and the seven LS calibration models necessary to determine the response (concentration of every drug) for every experiment of the design were estimated. The equations of these linear regression models “standardized peak area versus true concentration” figure in Table 4. A LS linear regression “predicted concentration versus true concentration” was also performed for every analyte in order to assess the trueness of the analytical method; the equations of these seven accuracy lines appear in the last column of Table 4. As the estimates of the intercept (b_0) and the slope (b_1) of a linear regression are correlated, the joint confidence region for both parameters of every accuracy model was estimated at $\alpha = 0.05$. The seven resulting ellipses are plotted together in Fig. 2. As can be seen there, the size of the seven ellipses was clearly different; this is due to the fact that the area of a joint confidence region is directly related to the residual variance of the model: the lower the value of the residual standard deviation (see Table 4, last column), the greater the quality of the precision on the predictions and the smaller the area within the confidence contour. The orientation of the ellipses for azaperone (in dark blue) and azaperol (in magenta), both standardized by the same IS (azaperone- d_4), differed from that of the confidence regions for the rest of the analytes. The reason was the use of a wider calibration range around the MRL of $100 \mu\text{g kg}^{-1}$ established in [23] for the sum of azaperone and azaperol, so higher values of the standard deviation of b_0 were obtained in both cases. Anyway, as the centroid of all the ellipses, represented by the least squares estimator (b_0, b_1), was statistically equal to (0, 1), it could be asserted that the trueness of the analytical method was ensured jointly for the seven drugs under study at a 95% confidence level.

The decision limit ($CC\alpha$) and the detection capability ($CC\beta$) of the analytical method for azaperone, azaperol and carazolol at its respective MRL with $\alpha = \beta = 0.05$ were estimated from the corresponding calibration model as it is laid down in [14]. The values of $CC\alpha$ and $CC\beta$ achieved for these three analytes when 1 replicate is performed were, respectively, $106.5 \mu\text{g kg}^{-1}$ and $112.5 \mu\text{g kg}^{-1}$ for azaperone, $104.4 \mu\text{g kg}^{-1}$ and $108.3 \mu\text{g kg}^{-1}$ for azaperol, and $6.12 \mu\text{g kg}^{-1}$ and $7.14 \mu\text{g kg}^{-1}$ for carazolol.

The eight runs that made up the experimental matrix together with the values of the seven responses (concentration of every drug) are gathered in Table 2. As the experimental design conducted was saturated, the eight coefficients β_i of each of the seven models could be exactly determined from Eq. (2); the values of β_1 to β_7 , related to the effect of the factors 1 to 7 on every response, are shown in Table 5. However, an estimate of the variance of the random term ε in Eq. (1) could not be achieved, so the statistical significance of neither the model nor that of its coefficients could be evaluated through the corresponding hypothesis test. For this work, the reproducibility standard deviation at a concentration c had been previously estimated for every analyte at an earlier stage of the validation process of the method. This standard deviation s , external to the experimental design, made it possible to determine whether the variations considered for the seven factors had a significant effect on every response thanks to the hypothesis test posed in Section 2.1.

For every analyte, the reproducibility standard deviation s at a specific concentration c , together with the value of the statistic β_i/s_{β_i} for the significance of each coefficient β_i ,

is shown in Table 5. For xylazine, propionylpromazine and haloperidol, the value of c corresponded to the final concentration of every analyte in a muscle sample fortified with 100 μL of the working standard solution of these sedatives; for chlorpromazine, which is a banned substance as tranquillizer in the European Union, and azaperone, azaperol and carazolol, with values of MRL established in porcine muscle [23], the value of c was exactly twice the decision limit ($CC\alpha$) of the method.

Taking the hypothesis test posed in Eq. (3) into account, there was not enough experimental evidence to reject the null hypothesis (H_0) in any case since no statistic value lay into the corresponding critical region (see Eq. (4)). Therefore, it could be concluded that the performance of the method was not influenced by the changes in the seven variables under study and the sample preparation procedure was robust.

Other alternatives for the evaluation of potentially influential factors when an estimation of the experimental variance is not used are Lenth's method and Bayesian analysis of the coefficients β_i , both of which were also applied in this work. These two strategies, which are explained in Section 2.2 and 2.3, respectively, are especially useful for the interpretation of saturated designs if an estimate of the residual error ε is not available.

The results of the application of Lenth's method are reflected in the last column of Table 5, where the confidence interval $(-t_{0.025,c/3} s_\beta, t_{0.025,c/3} s_\beta)$ estimated through this approach for every analyte is collected. It is clear that, in all cases, none of the seven coefficients β_1 to β_7 (also in Table 5) lay out of the corresponding interval, so all factors were non-significant and the sample preparation step was considered to be acceptably robust.

The second option applied for testing the significance of the effects was the Bayesian analysis of the coefficients of every model. As a result, the probabilities of each effect being active on every response (analyte) were calculated *a posteriori* for different values of k and α . Fig. 3 shows the range of *a posteriori* probabilities of significance plotted as a solid box between the maximum and minimum values of this probability. This range was determined for every analyte and every coefficient in Eq. (1) estimating the effect of a factor (β_1, \dots, β_7), so 49 ranges of *a posteriori* probabilities of significance are depicted in Fig. 3.

The maximum was below 60% in 35 cases (coefficients), between 64% and 80% in another 5, and higher than 80% for the remaining 9 cases. However, as the *a posteriori* probability of significance may really have any value between the minimum and the maximum, no factor gave clear proof of being active on the determination of none of the seven analytes. Only the way of performing the SPE step (factor 2) could be considered the most likely to be significant, since its *a posteriori* probability varied from nearly 45% to 95% in the case of azaperol, although it must be borne in mind that its real probability of significance might have any value within this range. As no effect had a minimum *a posteriori* probability of being significant close to 100%, no evidence of clearly influencing factors was provided, so the robustness of the sample preparation step could not be rejected. This conclusion exactly coincides with that drawn from each of the other two methodologies employed for the interpretation of the results of the Youden experimental design performed in this work (hypothesis testing and Lenth's method).

5. Conclusions

The robustness of the sample preparation stage for the determination of xylazine, azaperone, propionylpromazine, chlorpromazine, haloperidol, azaperol and carazolol in muscle tissue by LC/MS-MS has been verified by means of a 2_{III}^{7-4} fractional factorial

design, following the procedure recommended by Commission Decision No 2002/657/EC. After carrying out the eight runs that make up the experimental plan, the influence of seven factors involved in the sample preparation prior to analysis (namely, analyst, way of performance of the SPE step, age of both the phosphate buffer and eluent, volumes needed of the buffer, SPE washing solution and SPE eluent) has been evaluated by means of different statistical strategies, such as hypothesis testing, Lenth's method and Bayes' analysis, the latter two allowing to draw conclusions on which factors are active even though no estimate of the residual error is available. Whatever the approach used to interpret the results of the experimental design, the sample preparation procedure under study has proved to be robust to the small variations considered, which can be expected throughout its routine performance.

Acknowledgments

The authors thank the financial support provided by Ministerio de Ciencia e Innovación (CTQ2011-26022) and MINECO (CTQ2014-53157-R). M.L. Oca and L. Rubio are particularly grateful to University of Burgos for their FPI grants.

References

- [1] B. Magnusson, U. Ornemark (Eds.), Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, 2nd ed., 2014. ISBN 978-91-87461-59-0.
- [2] M. Thompson, S.L.R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC technical report), *Pure Appl. Chem.* 74 (2002) 835–855.
- [3] Y. Vander Heyden, A. Nijhuis, J. Smeyers-Verbeke, B.G.M. Vandeginste, D.L. Massart, Guidance for robustness/ruggedness tests in method validation, *J. Pharm. Biomed. Anal.* 24 (2001) 723–753.
- [4] Y. Vander Heyden, M. Jimidar, E. Hund, N. Niemeijer, R. Peeters, J. Smeyers-Verbeke, D.L. Massart, J. Hoogmartens, Determination of system suitability limits with a robustness test, *J. Chromatogr. A* 845 (1999) 145–154.
- [5] I. Taverniers, M. De Loose, E. Van Bockstaele, Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance, *Trends Anal. Chem.* 23 (2004) 535–552.
- [6] E. Karageorgou, V. Samanidou, Youden test application in robustness assays during method validation (Review), *J. Chromatogr. A* 1353 (2014) 131–139.
- [7] E. Aguilera, R. Lucena, S. Cárdenas, M. Valcárcel, E. Trullols, I. Ruisánchez, Robustness in qualitative analysis: a practical approach, *Trends Anal. Chem.* 25 (2006) 621–627.
- [8] L. Cuadros-Rodríguez, R. Romero, J. M. Bosque-Sendra, The role of the robustness/ruggedness and inertia studies in research and development of analytical processes, *Crit. Rev. Anal. Chem.* 35 (2005) 57–69.
- [9] ICH harmonised tripartite guideline – Validation of analytical procedures: text and methodology, Q2 (R1), 2005.
- [10] The United States Pharmacopeia 23/The National Formulary 18, United States Pharmacopeial Convention Inc., Rockville, 1995.

- [11] D.T. Burns, K. Danzer, A. Townshend, A tutorial discussion of the use of the terms "robust" and "rugged" and the associated characteristics of "robustness" and "ruggedness" as used in descriptions of analytical procedures, *J. Assoc. Public Anal.* 37 (2009) 40–60.
- [12] R.L. Plackett, J.P. Burman, The design of optimum multifactorial experiments, *Biometrika* 33 (1946) 305–325.
- [13] W.J. Youden, E.H. Steiner, *Statistical Manual of the AOAC – Association of the Official Analytical Chemists*, AOAC-I, Washington, D.C., 1975.
- [14] Commission Decision (EC) No 2002/657/EC of 12 August 2002, implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, *Off. J. Eur. Commun. L* 221 (2002) 8–36.
- [15] I. García, L. Sarabia, M.C. Ortiz, J.M. Aldama, Robustness of the extraction step when parallel factor analysis (PARAFAC) is used to quantify sulfonamides in kidney by high performance liquid chromatography-diode array detection (HPLC-DAD), *Analyst* 129 (2004) 766–771.
- [16] V. Gamba, C. Terzano, L. Fioroni, S. Moretti, G. Dusia, R. Galarini, Development and validation of a confirmatory method for the determination of sulphonamides in milk by liquid chromatography with diode array detection, *Anal. Chim. Acta* 637 (2009) 18–23.
- [17] N. Rodríguez, M.C. Ortiz, L.A. Sarabia, Study of robustness based on n-way models in the spectrofluorimetric determination of tetracyclines in milk when quenching exists, *Anal. Chim. Acta* 651 (2009) 149–158.
- [18] B. Solomun Kolanović, N. Bilandžić, I. Varenina, Validation of a multi-residue enzyme-linked immunosorbent assay for qualitative screening of corticosteroids in liver, urine and milk, *Food Addit Contam A* 28 (2011) 1175–1186.
- [19] M.R. Coleman, J.M. Rodewald, S.L. Brunelle, M. Nelson, L. Bailey, T.J. Burnett, Determination and confirmation of nicarbazin, measured as 4,4-dinitrocarbanilide (DNC), in chicken tissues by liquid chromatography with tandem mass spectrometry: First Action 2013.07., *J. AOAC Int.* 97 (2014) 630–640.
- [20] G.A. Lewis, D. Mathieu, R. Phan-Tan-Luu, *Pharmaceutical Experimental Design*, Marcel Dekker, Inc., New York, 1999.
- [21] N.A. Botsoglou, D.J. Fletouris, *Drug Residues in Foods: Pharmacology, Food safety and Analysis*, Marcel Dekker, Inc. New York, 2001.
- [22] H.F. De Brabander, H. Noppe, K. Verheyden, J. Vanden Bussche, K. Wille, L. Okerman, L. Vanhaecke, W. Reybroeck, S. Ooghe, S. Croubels, Residue analysis: Future trends from a historical perspective (Review), *J. Chromatograph. A* 1216 (2009) 7964–7976.
- [23] Commission Regulation (EU) No 37/2010 of 22 December 2009 on pharmacologically active substances and their classification regarding maximum residue limits in foodstuffs of animal origin, *Off. J. Eur. Union L* 15 (2010) 1–72.
- [24] Regulation (EC) No 470/2009 of the European Parliament and of the Council of 6 May 2009 laying down Community procedures for the establishment of residue

- limits of pharmacologically active substances in foodstuffs of animal origin, Off. J. Eur. Commun. L 152 (2009) 11–22.
- [25] R. Romero González, J.L. Fernández Moreno, P. Plaza Bolaños, A. Garrido Frenich, J.L. Martínez Vidal, Using mass spectrometry as a tool for testing for toxic substances in foods: toward food safety, *Rev Esp Salud Pública* 81 (2007) 461–474.
- [26] D.C. Montgomery, *Design and Analysis of Experiments*, eighth ed., John Wiley & Sons, Inc., New York, 2012.
- [27] R. Cela, M. Claeys-Bruno, R. Phan-Tan-Luu, in: S.D. Brown, R. Tauler, B. Walczak (Eds.), *Comprehensive Chemometrics*, Elsevier, Amsterdam, 2009, pp. 251–300.
- [28] F. Yates, *The Design and Analysis of Factorial Experiments*, Imperial Bureau of Soil Science, Harpenden, 1937.
- [29] S. Bratinova, B. Raffael, C. Simoneau, *Guidelines for performance criteria and validation procedures of analytical methods used in controls of food contact materials (EUR 24105 EN – Joint Research Centre – Institute for Health and Consumer Protection)*, Office for Official Publications of the European Communities, Luxembourg, 2009.
- [30] D. Mathieu, J. Nony, R. Phan-Tan-Luu, *NemrodW (Versión 2007_03)*, L.P.R.A.I. Marseille, France, 2007.
- [31] *STATGRAPHICS Centurion XVI (Version 16.1.11)*, StatPoint Technologies, Inc., Herndon, VA, 2010.

FIGURE CAPTIONS

- Fig. 1** Chemical structures of the seven drugs analysed.
- Fig. 2** Joint confidence region for the intercept (β_0) and the slope (β_1) of the accuracy line estimated for every analyte at a 5% significance level. (a) Xylazine: red line; (b) Azaperone: dark blue line; (c) Propionylpromazine: green line; (d) Chlorpromazine: yellow line; (e) Haloperidol: black line; (f) Azaperol: magenta line; (g) Carazolol: light blue line. The mean vector (b_0, b_1) of every ellipse is represented by +.

Fig. 3 Graphic representation of the results obtained, for each analyte, from the Bayesian analysis of the statistical significance of the seven factors evaluated (α belongs to the interval $[0.10,0.40]$, while k does to $[5.0,15.0]$).

ACCEPTED MANUSCRIPT

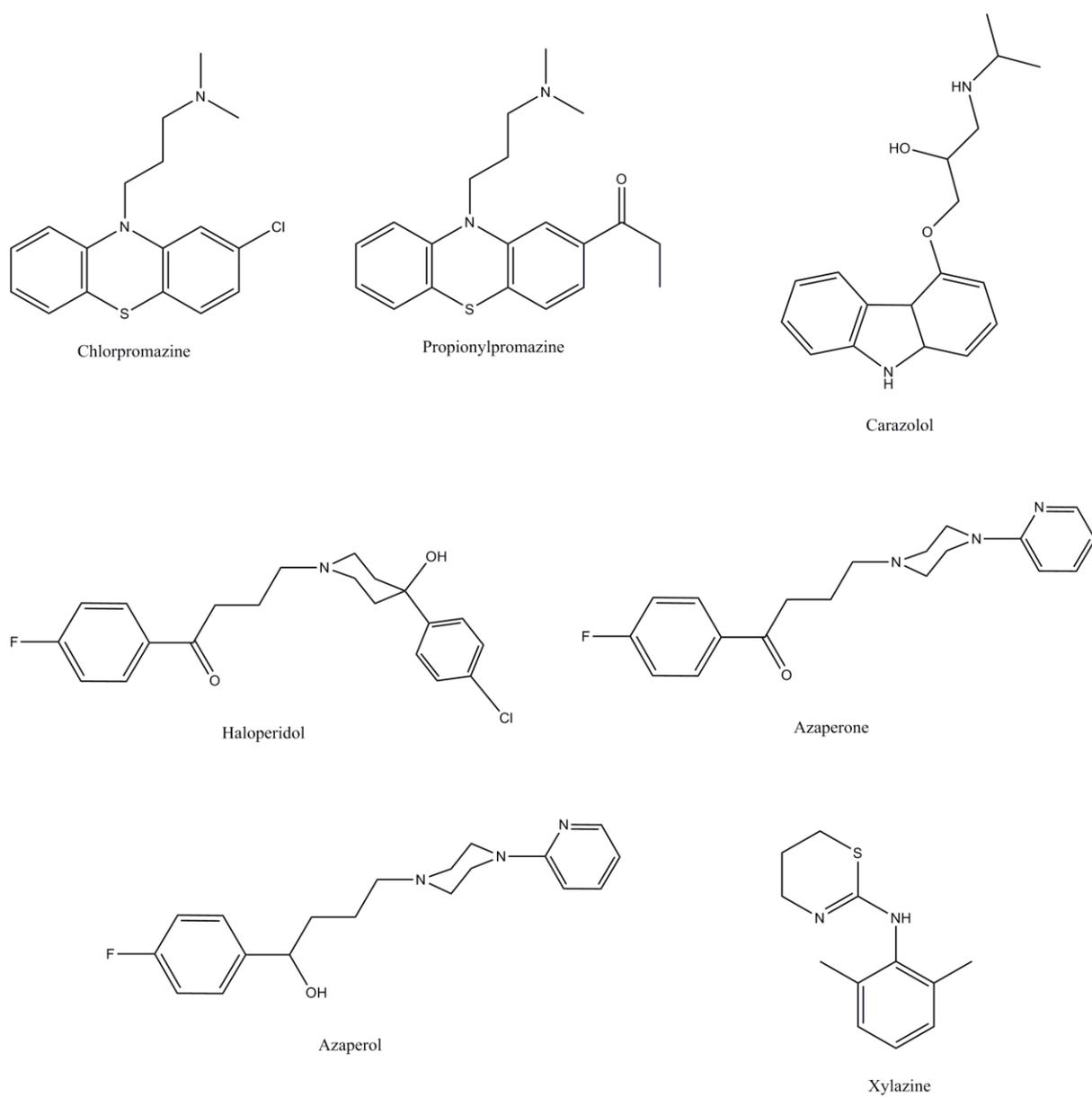


Fig. 1

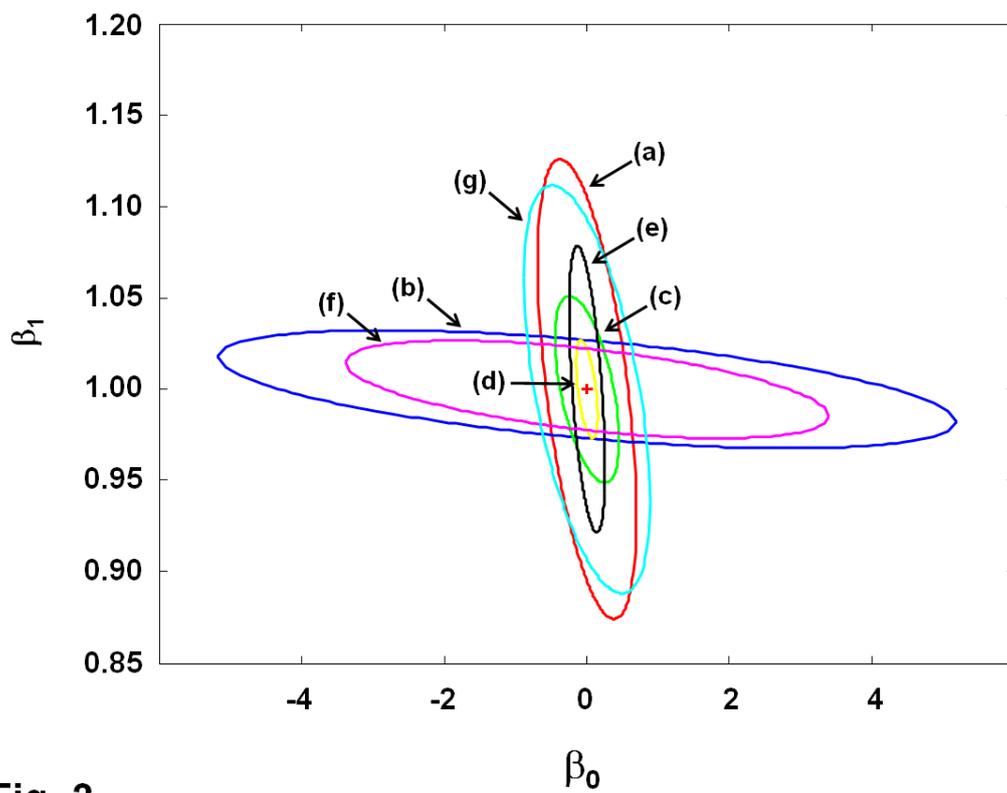


Fig. 2

ACCEPTED

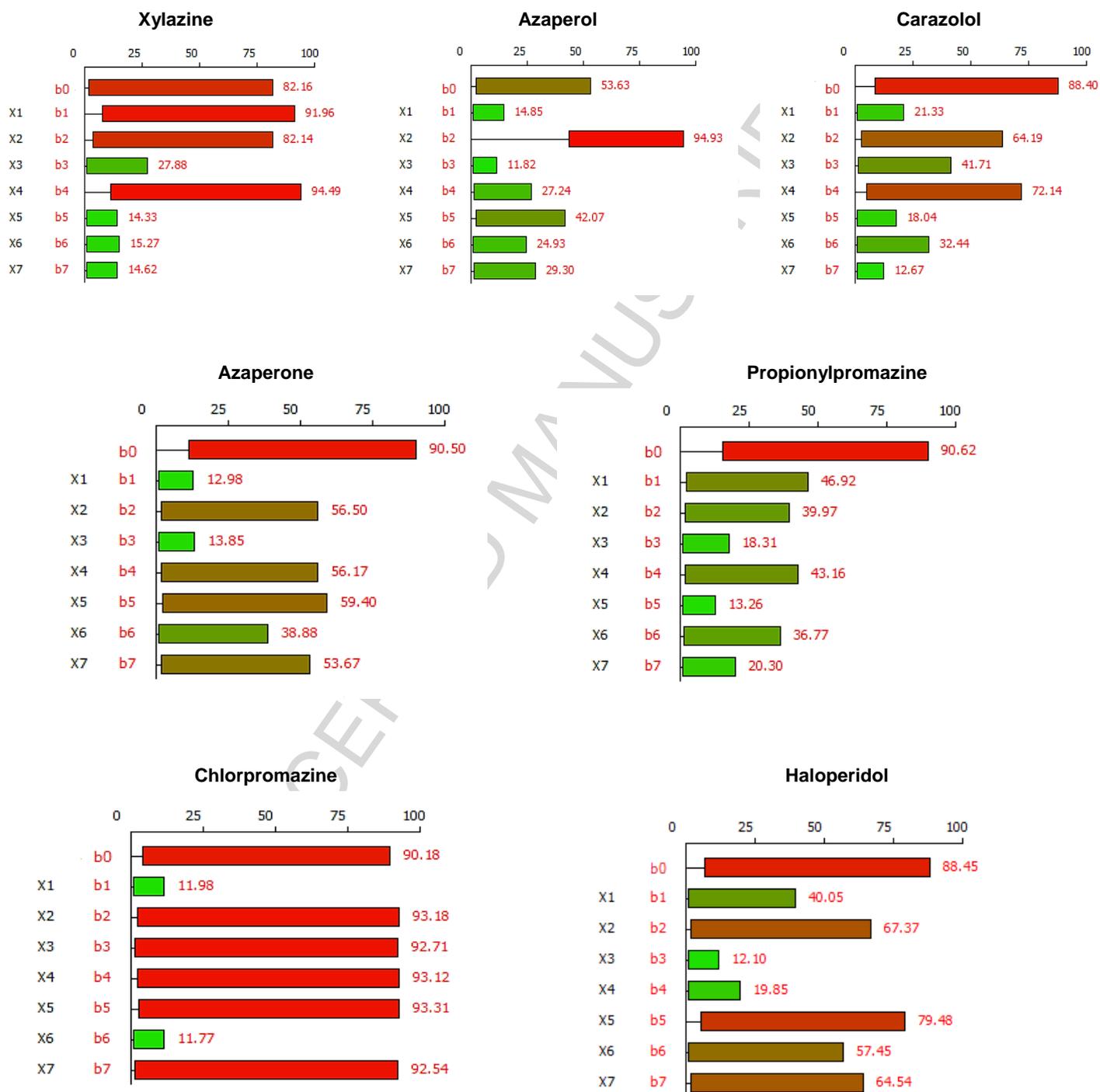


Fig. 3

Table 1

Experimental domain: factors and selected levels for the evaluation of the robustness of the

	Factor (Units)	Level	
		-1	+1
1	Analyst	Analyst A	Analyst B
2	SPE performance	Manual	Automated
3	Age of eluting solution	Newly prepared	1 week
4	Age of phosphate buffer	Newly prepared	1 month
5	Volume of phosphate buffer (mL)	10	9
6	Volume of eluting solution (mL)	4	4.5
7	Volume of washing solution (mL)	3	3.5

Table 2

Experimental matrix (in codified variables) for the robustness test and estimated concentrations of the seven analytes at each experiment.

Run	Level							Concentration ($\mu\text{g kg}^{-1}$)						
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	XYL	AZA	PROP	CHLOR	HALO	AZOL	CAR
1	-1	+1	-1	-1	+1	-1	+1	2.18	49.70	2.56	1.73	0.88	41.34	2.80
2	+1	+1	-1	+1	-1	-1	-1	2.13	51.13	2.20	1.81	1.04	42.09	2.46
3	-1	-1	-1	+1	+1	+1	-1	1.74	48.97	2.80	1.55	0.92	39.18	2.32
4	+1	+1	+1	+1	+1	+1	+1	2.04	49.45	2.40	1.69	0.95	40.86	2.90
5	-1	+1	+1	-1	-1	+1	-1	2.09	53.02	2.85	2.03	1.05	43.91	3.00
6	+1	-1	-1	-1	-1	+1	+1	2.21	50.57	2.65	1.73	0.96	39.84	2.69
7	-1	-1	+1	+1	-1	-1	+1	1.65	48.77	2.58	1.70	0.91	38.65	2.30
8	+1	-1	+1	-1	+1	-1	-1	2.25	49.99	2.69	1.76	0.90	38.81	2.77

Table 3

Transitions monitored for the determination of the seven analytes and the three internal standards by LC/MS-MS.

Compound	Most intense transition	Less intense transition
Xylazine	221.2 > 90.0	221.2 > 164.1
Azaperone	328.1 > 165.0	328.1 > 121.0

Propionylpromazine	341.2 > 86.1	341.2 > 58.1
Chlorpromazine	319.1 > 86.2	319.1 > 58.1
Haloperidol	376.2 > 165.1	376.2 > 123.2
Azaperol	330.3 > 121.1	330.3 > 149.2
Carazolol	299.2 > 116.1	299.2 > 222.1
Azaperone-d ₄	332.0 > 169.0	-
Chlorpromazine-d ₃	322.2 > 61.0	-
Haloperidol-d ₄	380.0 > 169.0	-

Table 4

Calibration models for the quantification of the seven drugs analysed.

Analyte	Calibration range ^a ($\mu\text{g kg}^{-1}$)	$A_{\text{analyte}}/A_{\text{IS}}^b = f(C_{\text{true}})$ (R^2 ; s_{yx})	$C_{\text{pred}} = b_0 + b_1 C_{\text{true}}$ (R^2 ; s_{yx})
XYL	0 – 6.92	$y = 1.04 + 4.68x$ (98.41%; 1.77)	$y = 5.6 \cdot 10^{-5} + 1.00x$ (98.41%; 0.38)
AZA	0 – 202.00	$y = 0.09 + 0.06x$ (99.89%; 0.17)	$y = -1.3 \cdot 10^{-5} + 1.00x$ (99.89%; 2.84)
PROP	0 – 11.15	$y = -0.64 + 2.37x$ (99.73%; 0.59)	$y = -5.6 \cdot 10^{-6} + 1.00x$ (99.73%; 0.25)
CHLOR	0 – 7.26	$y = -0.06 + 0.67x$ (99.93%; 0.06)	$y = 1.5 \cdot 10^{-5} + 1.00x$ (99.93%; 0.08)
HALO	0 – 4.00	$y = -2.5 \cdot 10^{-3} + 3.99x$ (99.38%; 0.54)	$y = 2.9 \cdot 10^{-5} + 1.00x$ (99.38%; 0.14)
AZOL	0 – 160.00	$y = 4.9 \cdot 10^{-3} + 0.07x$ (99.93%; 0.13)	$y = -1.5 \cdot 10^{-4} + 1.00x$ (99.93%; 1.86)
CAR	0 – 10.03	$y = 0.08 + 2.61x$ (98.75%; 1.27)	$y = 7.4 \cdot 10^{-5} + 1.00x$ (98.75%; 0.49)

^(a) n = 6 calibration standards^(b) Standardized peak area of the analyte, calculated as Absolute peak area of the analyte/Absolute peak area of its internal standard

Table 5
Evaluation of the significance of the effects of the seven factors from a hypothesis test using an external variance and from Lenth's method.

Analyte	External reproducibility variance										Lenth's method	
	c ($\mu\text{g kg}^{-1}$)	s ($\mu\text{g kg}^{-1}$)	Coefficients and statistic	Factor								Confidence interval
				1	2	3	4	5	6	7		
XYL	1.73	0.36	β_1	0.12	0.07	-0.03	-0.15	0.02	-0.02	-0.02	(-0.25, 0.25)	
			β_1/s_R	0.96	0.57	-0.23	-1.13	0.12	-0.14	-0.13		
AZA	50.50	2.02	β_1	0.09	0.62	0.11	-0.62	-0.67	0.30	-0.58	(-3.40, 3.40)	
			β_1/s_R	0.12	0.87	0.15	-0.87	-0.94	0.42	-0.81		
PROP	2.79	0.42	β_1	-0.10	-0.09	0.04	-0.10	0.02	0.08	-0.04	(-0.49, 0.49)	
			β_1/s_R	-0.70	-0.61	0.26	-0.65	0.14	0.56	-0.30		
CHLOR	1.81	0.18	β_1	-0.003	0.06	0.05	-0.06	-0.07	0.001	-0.04	(-0.27, 0.27)	
			β_1/s_R	-0.05	1.00	0.72	-0.97	-1.05	0.01	-0.58		
HALO	1.00	0.07	β_1	0.01	0.03	0.002	0.005	-0.04	0.02	-0.03	(-0.11, 0.11)	
			β_1/s_R	0.39	1.12	0.08	0.19	-1.53	0.75	-1.04		
AZOL	40.00	4.40	β_1	-0.19	1.47	-0.03	-0.39	0.54	0.36	-0.41	(-2.44, 2.44)	
			β_1/s_R	-0.12	0.94	-0.02	-0.25	-0.34	0.23	-0.27		
CAR	2.51	0.38	β_1	0.05	0.14	0.09	-0.16	0.04	0.07	0.02	(-0.41, 0.41)	
			β_1/s_R	0.36	1.03	0.66	-1.20	0.30	0.53	0.13		

ACCEPTED

HIGHLIGHTS

Robustness study of the sample preparation step to determine sedatives by LC/MS-MS

Seven factors evaluated through a Youden design according to the European legislation

Results interpreted from quite useful methodologies when the residual error is unknown

Lenth's and Bayes' approaches for robustness testing when performing saturated designs