When LC-HRMS metabolomics gets ISO17025 accredited and ready for official controls - Application to the screening of forbidden compounds in livestock

Gaud Dervilly-Pinel, Anne-Lise Royer, E. Bozetta, M. Pezzolato, Loïc Herpin, S. Prevost, Bruno Le Bizec

When LC-HRMS metabolomics gets ISO17025 accredited and ready for official controls – Application to the screening of forbidden compounds in livestock

Gaud DERVILLY-PINEL[1]*, Anne-Lise ROYER[1], Elena BOZETTA[2], Marzia PEZZOLATO[2], Loïc HERPIN[1], Stéphanie PREVOST[1], Bruno LE BIZEC[1]

[1]Laberca, Oniris, INRA, Université Bretagne Loire, 44307, Nantes-France

[2] Istituto Zooprofilattico Sperimentale del Piemonte Liguria e Valle d'Aosta, 10154 Torino - Italia

*corresponding author

gaud.dervilly@oniris-nantes.fr

laberca@oniris-nantes.fr

ABSTRACT

Within the particular context of controlling chemical residues in food, an alternative to targeted approaches has emerged; it consists in the characterization of physiological perturbations induced upon exposure of animals to a given chemical substance/class of substances to highlight suitable biomarkers addressing safety and/or regulatory issues. Metabolomics in particular has been investigated in the hope of identifying such biomarkers, and a range of studies have, from that time forward, demonstrated the efficiency of the strategy.

Until very recently, steps toward official or commercial implementation of corresponding tools had however still to be taken. In particular, the lack of guidelines and criteria to validate such methods that do not target specific chemical species *per se*, constituted a bottleneck.

In the present work, a metabolomics model dedicated to the detection of β-agonists administration in bovine has been developed and fully validated. Validation criteria (selectivity, robustness, stability, suspicion threshold definition, false positive and false negative rates) have been proposed in agreement with EU expectations (Dec 2002/657), enabling demonstrating performances complying screening requirements.

Although some of the biomarkers involved in the prediction model remain still un-elucidated, the corresponding LC-HRMS method has recently been ISO17025 accredited, allowing for the very first official implementation of a metabolomics based strategy within French National Monitoring Plans.

KEY WORDS: metabolomics, validation, performance criteria, β-agonists, calves, urine, chemical food safety, public health, risk management

INTRODUCTION

In order to meet current demands from official laboratories, consumers and regulatory agencies on food safety, quality and traceability, in the globalised 21st century, it is now even more necessary for official laboratories to develop faster, powerful, cleaner, cheaper and well-characterised analytical methodologies. These methods should ideally be both robust and reproducible as well as being capable of providing information about food adulteration, contamination, traceability, etc., while ensuring compliance with food and trade laws. Food quality has indeed become a major issue worldwide and there is a strong consumer demand to ensure both safety (by detecting the presence of (chemical or microbiological) toxins, contaminants or residues of chemical substances) and authenticity (geographic as well as potentially botanical origin, mislabelling, adulteration...). Generally, efficient targeted methodologies are effective in addressing the main potential issues (Wang, Wang, & Cai, 2013). However, in some cases these targeted approaches do not allow the detection of emerging illicit practices and therefore new approaches and strategies are demanded. Such a need for new and alternative screening strategies to detect fraudulent practices impacting on food quality, which are based on non-targeted analysis, have thus gained attention of both the research and the risk assessors community over the last past decade (Chassy, 2010; Davies, 2010; Pielaat, et al., 2013).

Thereby, alternative strategies aiming at identifying specific profiles or biomarkers, associated with given food quality issues, have emerged through the development of high-throughput, non-targeted and broad scale approaches to analyse metabolite content to provide fingerprints which can be used for characterisation of authenticity and quality (Castro-Puyana & Herrero, 2013; Ellis, et al., 2012; Milgrom & Baulieu, 2008; Oms-Oliu, Odriozola-Serrano, & Martín-Belloso, 2013). Metabolomics based on NMR spectroscopy or mass spectrometry (most often linked with chromatography) in particular has been investigated in the hope of identifying new food quality biomarkers (Cubero-Leon, Peñalver, & Maquet, 2014; G. Dervilly-Pinel, et al., 2012; Gallart Ayala, Chéreau, Dervilly-Pinel, & Le Bizec, 2015). From that time forward, several studies have demonstrated the efficiency of untargeted fingerprinting (Clara Ibáñez, García-Cañas, Valdés, & Simó, 2013; Clara Ibáñez, Simó, García-Cañas, Cifuentes, & Castro-Puyana, 2013; C. Ibáñez, et al., 2012). Although this research is a mandatory step in initial assessment of the strategy, steps toward official or commercial implementation of corresponding (screening) tools are still to be taken (Esslinger, Riedl, & Fauhl-Hassek, 2014; Riedl, Esslinger, & Fauhl-Hassek, 2015). As these methods do not target specific chemical species per se there is in particular a need to harmonise and establish internationally accepted validation and standardisation protocols to be applied to such methods (Creek, et al., 2014; Gika, Theodoridis, Plumb, & Wilson, 2014; Goodacre, et al., 2007; Luque de Castro & Priego-Capote, 2017; Martin, et al., 2015; MSI, et al., 2007; Rocca-Serra, et al., 2016). These have to be thoroughly developed and shared with actors implicated in the field

(researchers, control laboratories, risk assessors, policy makers...) in order to enable universal acceptance.

Although many guidelines for validating targeted methods are available and provide actors with performance indicators (e.g. pesticides (SANCO 12495/2011) (EuropeanCommission, 2011), pharmacological drugs (Dec 2002/657/EC) [23] (Council Decision 2002/657/EC, 2002)), such equivalent guidelines for untargeted method validation are not available yet, which certainly explains their limited application for routine or official control purposes. The main difference between non-targeted and targeted approaches is related to the multivariate nature of fingerprints in the former case, compared to single or multi-chemical species analysis in the latter one. In targeted strategies, results and method performance are evaluated compound-by-compound using univariate statistics, while multivariate analysis (which are generally referred to chemometric methods) are required to evaluate relevance of patterns arising from fingerprinting approaches (Alonso, Marsal, & Julia, 2015; Godzien, Alonso-Herranz, Barbas, & Grace Armitage, 2015). Guidelines for biomarkers/models discovery and validation are therefore urgently required by the scientific community (Dudzik, Barbas-Bernardos, Garcia, & Barbas, 2017; Goodacre, et al., 2007; Naz, Vallejo, García, & Barbas, 2014).

Validation of untargeted approaches has to be defined at various steps (Dudzik, et al., 2017) corresponding to analytical, statistical and biological stages. Consequently different validation levels may be defined (Riedl, et al., 2015), the first one being related to the initial model establishment and validation; it refers to the R&D work (models generation, verification and validation) (Level 1 in Figure 1), while the second one corresponds to the validation of the resulting method (Level 2 in Figure 1), therefore semantics has to be clearly defined within these two levels since similar wordings would refer to different steps, skills, tools…. The definition and measurement of performance indicators have to be defined at these two levels, since corresponding and respective requirements and expectations are different. In this context, clarifying the different steps and levels to be considered is a priority and it appears for instance that different validation steps will have to be performed, and that both performance characteristics and criteria will have to be specifically defined at the R&D level and also at the validation level.

Regarding the potential of such untargeted tools and related biomarkers in controlling fraudulent practices in food producing animals provides the following picture. Issues encountered in controlling the use of growth promoters in livestock's led over the last past 10 years to the investigation of the physiological effects induced as a consequence of illegal practices. With the objective of revealing biomarkers of exposure that may subsequently be monitored for screening purposes omics strategies have already proven their relevance using respectively either transcriptomics (Reiter, Walf, Christians,

Pfaffl, & Meyer, 2007; Riedmaier, Tichopad, Reiter, Pfaffl, & Meyer, 2009), proteomics (Cacciatore, et al., 2009; Kinkead, Elliott, Cannizzo, Biolatti, & Mooney, 2015; Nebbia, et al., 2011), or metabolomics (Courant, et al., 2009; G. Dervilly-Pinel, et al., 2012; C. Jacob, Dervilly-Pinel, Biancotto, Monteau, & Le Bizec, 2015; C. C. Jacob, Dervilly-Pinel, Biancotto, & Le Bizec, 2013; Kouassi Nzoughet, Gallart-Ayala, Dervilly-Pinel, Biancotto, & Le Bizec, 2015; J. J. Kouassi Nzoughet, et al., 2015; Laparre, et al., 2017; Pinel, et al., 2010; Pinel, et al., 2011).

Compared to other omics strategies, metabolomics gathered more interest from the residue-control world since it involves analytical platforms similar to those already available in research laboratories in this area. Furthermore, biomarkers evidenced upon metabolomics studies are considered as easier to subsequently monitor since involving targeted analytical strategies close to those already available in corresponding laboratories in charge of the control.

Based on a metabolomics model dedicated to the detection of β-agonists administration in bovine, which validation (Level 1) has previously been published (Gaud Dervilly-Pinel, Chereau, Cesbron, Monteau, & Le Bizec, 2015), the present article describes level 2 validation (Figure 1) and proposes performance criteria (selectivity, robustness, stability, suspicion threshold definition, false positive and false negative rates) in agreement with EU expectations (Dec 2002/657) in the objective of validating a screening method. Further, as discussion, we describe the process toward ISO17025 accreditation as well as practical aspects in relation with routine implementation of the method (Level 3, Figure 1).

1- MATERIALS & METHODS

1.1 Biological samples

Urine samples have been collected in the frame of several animal experiments as described in Table 1. In addition, a large number (> 200) of samples arising from the 2013 French monitoring plans representative of the national production have been including to ensure significant variability to be considered within the developed models (Table 2). These urine samples which have been previously characterized and declared compliant, have been used as control samples in the present study.

1.2 Metabolomics general workflow

Urine samples are filtered over 10 KDa membranes under centrifugation at 10,000 g at 5 °C for 20 minutes. Chromatographic separation is performed with a high performance liquid chromatography (HPLC) system (1200 Infinity Series from Agilent Technologies, Santa Clara, California, USA) on a Hypersil Gold C18 column (2.1 mm x 100 mm, 1.9 µm particle size, Thermo Fisher Scientific). Mobile phase consists in water containing 0.1% acetic acid (A) and acetonitrile containing 0.1% acetic acid (B). Liquid Chromatography - High Resolution mass spectrometry (LC-HRMS) fingerprints are acquired

either on an LTQ-Orbitrap/ Exactive or Q-exactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) in positive heated electrospray ionization mode. Full scan mass spectra are acquired from m/z 80 to m/z 1000 using a mass resolution of 50 000 FWHM at 400 m/z in centroid mode. Intensities of the three biomarkers of interest M1, M2 and M3 are reported within the following mathematical equation validated in previous work (Gaud Dervilly-Pinel, et al., 2015).

$$Y = -0.91.(\text{Intensity M1}) + 0.44.(\text{Intensity M2}) + 0.37.(\text{Intensity M3})$$

The Y value thus obtained corresponds to the prediction value (arbitrary unit) of the position of the urine sample tested on the model.

2- RESULTS & DISCUSSION

2.1 Suspicious threshold definition

Validating the method (Level 2, Figure 1) and assessing its screening performances required setting a decision threshold, which was defined as follows. The threshold has been determined on the basis of 207 compliant urine samples collected in the frame of Exp A (n=147) and Exp. B (n=60). These control samples were obtained from both control animals and animals from the treated groups before clenbuterol administration. The 207 analyzed samples leaded to a mean prediction value $\mu_{control}$ of -2.27 with a standard deviation σ of 0.36. The distribution of the prediction values is presented on Figure 2 and attests for a normal distribution. A decision threshold (T), enabling concluding on sample compliance, has consequently been set and calculated as the mean value ($\mu_{control}$) of all these control samples added to 2 times the standard deviation of the same samples ($T = \mu_{control} + 2\sigma$) which corresponds to the 95th percentile level of confidence. Decision threshold value has consequently been set at T=-1.55. Such a strategy has already been successfully proposed (Gaud Dervilly-Pinel, et al., 2015; Kaabia, et al., 2014).

2.2 Applicability of the method

Applicability of the method was assessed with regard to threshold defined above.

2.2.1 Selectivity toward β-agonists treatments

While the model has already been successfully challenged for the prediction of different β-agonists practices (various drugs, doses, mixtures, treatment length) within Exp. A to D (Table 1) (Gaud Dervilly-Pinel, et al., 2015), selectivity of the method was further assessed on the basis of urine samples collected on 12 additional animals (8 clenbuterol-treated, 4 controls) which were of a different breed from previous animals involved and were further raised under different breeding conditions in Italy (Exp. E). Urine samples collected on control animals during the duration of the experiment (n=65) were all predicted as compliant with a mean value of -2.36 (σ = 0.64). Figure 3 illustrate the prediction of

urine samples from the treated group over the period of the experiment. While three samples are mi-classified before the administration starts (false suspicious), all samples (n=40) were plotted above the threshold (i.e. suspicious) from the first day and until 25 days after clenbuterol administration starts, i.e. 5 days after the end of the treatment. Some animals (e.g. calf 8) even positively responded as soon as 8 hours after the beginning of the treatment. Some particular animals (calves 1, 3 and 5) remained even suspicious for a longer period after the end of the treatment (Figure 3).

Such results combined with those previously reported for Exp. A to D (Gaud Dervilly-Pinel, et al., 2015) allows extending the application of the tool to a range of calves' breeds, whatever the breeding place and conditions, ensuring thus a potential for applicability of the newly developed screening tool.

2.2.2 Selectivity toward general bovine population

Subsequently, a large number (> 200) of urine samples collected from animals' representative of the French production have been including in the validation to ensure significant variability to be considered within the prediction method (Table 2). These urine samples have previously been characterized and declared compliant. The objective was to refine method's selectivity and define more precisely the scope of application of such a screening strategy which initial models have been established only based on young male bovine animals. In particular, main purpose was to evaluate prediction ability toward bovine adults. The 201 urine samples have been analysed according to the developed workflow and predicted on the established model. Considering the threshold value defined above (§ 2.1), prediction resulted in 87 samples (43.3 % of the general population studied) above the defined threshold (Figure 4a), which is not acceptable. Investigating further the animals' characteristics (age, gender, production type) leaded to consider only animals below 6 months old as relevant target population for the strategy. Considering this calve population, 20% (14/71) was predicted with a value above the proposed threshold (Figure 4b).

2.2.3 Method performances

In the objective of developing a screening tool for control purposes, with a scope of application defined as follows: target compounds = β-agonists, population = calves (i.e. bovine < 6 months), matrix = urine, detection time window = up to 5 days after the end of treatment, results obtained in both the present study and the previous one (Gaud Dervilly-Pinel, et al., 2015) have been gathered to evaluate method performances considering the threshold $T = \mu_{control} + 2\sigma$. All urine samples ($\Sigma n=517$) arising from Exp. A to E (both control and treated population, n=446)) and from general calve population (n=71) have been included in this step. Performances (Table 3) correspond to a false suspicious rate (false positive) of 4.8% and false compliant score (false negative) of 1.5%, which fully falls within expected

performances for screening purposes as described in Dec 2002/657/EC where a false compliant rate below 5% is required.

2.2.4 Selectivity toward other veterinary drugs treatments

In the objective of further assessing method performances and biomarkers' behaviors toward other veterinary drugs treatments (forbidden or allowed), additional steps in the validation process have been included as follows:

Steroids. The developed method with proven relevance in highlighting β-agonists abuse in bovine has further been tested for its potential in signing steroids administration (Exp. F and G). Upon characterization of the urine samples from both experiments, a metabolic disruption of the three monitored biomarkers (M1, M2 and M3) could be observed for all steroid treated animals, thus confirming disruption of the same molecular targets. This suggests such biomarkers are not only specific of β-agonists treatment but may also be involved in steroids related effect. However, the present method cannot as such be extended to a screening tool including steroid illegal practices in bovines. At least the decision threshold should be adapted in that context. Indeed, while the application of the threshold criteria to the 40 urine samples from Exp. F and G leaded to false positive rate below 8%, the false negative rate was up to 50 % which does not fit with screening requirements.

Corticosteroids. Selectivity of the method also had to be checked with regard to other veterinary drugs that might be used in calves, and that are known to be associated with growth promoters' properties. In that respect, urine samples collected in the frame of a large corticosteroid experimental design involving both prednisolone and dexamethasone (Exp. H) have been predicted on the model. The complete set of samples consisted in 103 urines that were all (except 3) plotted below (μ=-2.67) the defined decision threshold (T=-1.55). Such result on the one hand indicates that corticosteroids do not affect the three markers profile, on the other one confirms method's performances since only 3.1% of the samples would be falsely declared as suspicious upon this screening test, which fully comply with EU expectations in that context.

2.3 Robustness

2.3.1 Repeatability and reproducibility of the measure

Repeatability of the prediction was assessed through the analysis of a same compliant sample 10 times within one single analytical run. As a result, a mean prediction value of -2.43 associated with a standard deviation of 0.04 was obtained, satisfying repeatability expectations.

Reproducibility (intra-lab) of the measure has been evaluated using control charts dedicated to the prediction of reference compliant (n=2) and suspicious (n=2) samples, analyzed over several analytical runs during a 6 months' period. During the study, samples were stored as aliquots at -20 °C. Associated

predictions on the model (Figure 5) enable demonstrating reproducibility of the measure. Further, such results also attest for the relevance of storage conditions.

Finally, instrumental reproducibility was evaluated. While a LTQ-Orbitrap system has been used for model development and validation (Level 1, Figure 1), as well as for method validation (Level 2, Figure 1) described above, the ability of both Exactive and Q-Exactive systems to monitor markers of interest, with intensities of the measures allowing to use the model, has been tested. Results obtained on both compliant and ~~non-compliant~~non-compliant samples prediction using both alternative instruments allowed concluding to a maximum 15% bias of the prediction, without affecting status prediction.

2.3.2 Markers and samples stability

Stability of the samples and associated markers has been assessed above (§ 2.3.1) through the analysis of compliant and ~~non-compliant~~non-compliant samples, aliquoted and stored at -20°C during 6 months.

With similar objective, two samples of urine (one compliant and one ~~non-compliant~~non-compliant) have been submitted to several (n=8) freeze-thawing cycles over two months and the samples analyzed at each thawing step. A decrease in the prediction values could be observed as a function of time, which, because reproducibility of the measure has previously been demonstrated, attested for biomarkers degradation upon several (>3) freeze-thawing cycles. However, it is important to note that status prediction remained unchanged, especially for the ~~non-compliant~~non-compliant sample, which could still be predicted as suspicious even with decreased prediction value. A maximum number of freeze-drying cycles was fixed at three.

Finally, samples stability was tested at room temperature. Four urine samples (2 compliant and 2 ~~non-compliant~~non-compliant) have been let for 48 hours at room temperature in the laboratory before analysis. All samples were correctly predicted on the model, attesting for markers stability in urine at room temperature over at least 2 days.

2.4 Routine implementation & Accreditation

While upon validation process described above, applicability of the method could be demonstrated together with performances in accordance with screening requirements, the method was subsequently candidate to routine implementation for control purposes (Level 3, Figure 1). In particular, main expectation consists here in transferring biomarkers monitoring to more routine dedicated instruments (e.g. triple quadrupole MS). A number of key challenges however has to be overcome to deliver such objective: i) the identification of specific fragmentation patterns to establish SRM transitions for targeted monitoring and ii) the adaptation of chromatographic conditions to both allow detection of the signal of interest in a known Rt range and to reduce the chromatographic

separation time to fulfil high sample throughput requirements. In the present study, main issue encountered related to the transfer step relied on the fact that two of the three involved biomarkers remains un-elucidated, which impairs classical targeted approach involving specific sample preparation and monitoring SRM transitions to be performed. In that context the untargeted LC-HRMS full scan approach used for biomarkers discovery and validation was subsequently transferred as such for routine purposes. Working "blindy" i.e. with signals only characterized with an exact mass and a chromatographic retention time (LC-HRMS technology) without the possibility to use reference standard compounds consequently involves the resort to a panel of precautions to be taken, such as the preparation of QC samples, the use of a set of pseudo-internal standards, and the characterization of both compliant and ~~non-compliant~~non-compliant reference samples all along the analytical sequence. In that context, a typical routine analytical run is composed as illustrated in Supplementary Figure S1.

Finally, examination of the screening method against ISO 17025 standards met expectations of the accreditation body which enabled the method to be accredited and implemented as such within French National Monitoring Plans.

CONCLUSION

The process described above enabled validating a method to screen for β-agonists in calves, based on a metabolomics approach. The validation process allowed describing method performances with a decision threshold set whatever the HRMS Orbitrap system used. Screening was demonstrated possible as soon as the first day after β-agonist treatment starts and efficient at least during 5 days following the end of the application, with performances of 1.5% false negative and 4.8% false positive rates. Such performances enabled the strategy to be transferred for control purposes within national monitoring plans and to be acknowledged ISO17025, although some of the biomarkers still remain unknown.

Next step will focus on the structural elucidation of the chemical signals highlighted as candidate biomarkers. Up to now, any attempts to elucidate these markers, using mass spectrometric (MS) strategies only, have failed. The identification of metabolites is however essential to convert analytical data into meaningful biological knowledge, and further validate markers relevance. Identification is overall a pre-requisite in the objective of transferring biomarkers monitoring from a research instruments (i.e. High Resolution MS) to more routine dedicated instruments (Triple quadripole MS for instance). A targeted method is indeed expected to provide greater selectivity and sensitivity than untargeted one. Such approach would involve quantification of the metabolites through the use of internal standards and authentic chemical standards to construct calibration curves for each of the

markers. In that context, the question that may arise is the following “does it make sense to validate an untargeted metabolomics approach?” since as soon as the biomarkers are unambiguously/tentatively identified, the strategy is expected to move from untargeted metabolomics to a targeted accredited method… In the present case, having accredited the untargeted workflow was the only possible way to officially implement such innovative screening tool since identification of 2 markers still remains unresolved.

References

Alonso, A., Marsal, S., & Julia, A. (2015). Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol, 3*, 23.

Cacciatore, G., Eisenberg, S. W., Situ, C., Mooney, M. H., Delahaut, P., Klarenbeek, S., Huet, A. C., Bergwerff, A. A., & Elliott, C. T. (2009). Effect of growth-promoting 17beta-estradiol, 19-nortestosterone and dexamethasone on circulating levels of nine potential biomarker candidates in veal calves. *Anal Chim Acta, 637*(1-2), 351-359.

Castro-Puyana, M., & Herrero, M. (2013). Metabolomics approaches based on mass spectrometry for food safety, quality and traceability. *TrAC Trends in Analytical Chemistry, 52*(0), 74-87.

Chassy, B. M. (2010). Can –omics inform a food safety assessment? *Regulatory Toxicology and Pharmacology, 58*(3, Supplement), S62-S70.

Council Decision 2002/657/EC. (2002). Council Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, . *Off J Eur Union*.

Courant, F., Pinel, G., Bichon, E., Monteau, F., Antignac, J. P., & Le Bizec, B. (2009). Development of a metabolomic approach based on liquid chromatography-high resolution mass spectrometry to screen for clenbuterol abuse in calves. *Analyst, 134*(8), 1637-1646.

Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E. L., Sumner, L. W., Trengove, R., & Wolfender, J.-L. (2014). Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics, 10*(3), 350-353.

Cubero-Leon, E., Peñalver, R., & Maquet, A. (2014). Review on metabolomics for food authentication. *Food Research International, 60*(0), 95-107.

Davies, H. (2010). A role for "omics" technologies in food safety assessment. *Food Control, 21*(12), 1601-1610.

Dervilly-Pinel, G., Chereau, S., Cesbron, N., Monteau, F., & Le Bizec, B. (2015). LC-HRMS based metabolomics screening model to detect various β-agonists treatments in bovines. *Metabolomics, 11*, 403-411.

Dervilly-Pinel, G., Courant, F., Chereau, S., Royer, A., Boyard-Kieken, F., Antignac, J., & Le Bizec, B. (2012). Metabolomics in food analysis: application to the control of forbidden substances. *Drug Test Anal, 4*, 10.

Dudzik, D., Barbas-Bernardos, C., Garcia, A., & Barbas, C. (2017). Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J Pharm Biomed Anal*.

Ellis, D. I., Brewster, V. L., Dunn, W. B., Allwood, J. W., Golovanov, A. P., & Goodacre, R. (2012). Fingerprinting food: current technologies for the detection of food adulteration and contamination. *Chem. Soc.rev., 41*, 21.

Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International, 60*, 189-204.

EuropeanCommission. (2011). Method validation and quality control procedures for pesticide residues analysis in food and feed. In *Document N° SANCO* (Vol. 12495).

Gallart Ayala, H., Chéreau, S., Dervilly-Pinel, G., & Le Bizec, B. (2015). Potential of mass spectrometry metabolomics for chemical food safety. *Bioanalysis Review, 7*(1), 133-146.

Gika, H. G., Theodoridis, G. A., Plumb, R. S., & Wilson, I. D. (2014). Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J Pharm Biomed Anal, 87*(0), 12-25.

Godzien, J., Alonso-Herranz, V., Barbas, C., & Grace Armitage, E. (2015). Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample. *Metabolomics, 11*, 11.

Goodacre, R., Broadhurst, D., Smilde, A., Kristal, B., Baker, J., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A., Ebbels, T., Kell, D., anetti, C., Newton, J., Paternostro, G., Somorjai, R., Sjöström, M., Trygg, J., & Wulfert, F. (2007). Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics, 3*, 11.

Ibáñez, C., García-Cañas, V., Valdés, A., & Simó, C. (2013). Novel MS-based approaches and applications in food metabolomics. *TrAC Trends in Analytical Chemistry, 52*, 100-111.

Ibáñez, C., Simó, C., García-Cañas, V., Cifuentes, A., & Castro-Puyana, M. (2013). Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics: A review. *Anal Chim Acta, 802*(0), 1-13.

Ibáñez, C., Valdes, A., Garcia-Canas, V., Simo, C., Celebier, M., Rocamora-Reverte, L., Gomez-Martinez, A., Herrero, M., Castro-Puyana, M., Segura-Carretero, A., Ibanez, E., Ferragut, J. A., & Cifuentes, A. (2012). Global Foodomics strategy to investigate the health benefits of dietary constituents. *J Chromatogr A, 1248*, 139-153.

Jacob, C., Dervilly-Pinel, G., Biancotto, G., Monteau, F., & Le Bizec, B. (2015). Global urine fingerprinting by LC-ESI(+)-HRMS for better characterization of metabolic pathway disruption upon anabolic practices in bovine. *Metabolomics, 11*, 184-197.

Jacob, C. C., Dervilly-Pinel, G., Biancotto, G., & Le Bizec, B. (2013). Evaluation of specific gravity as normalization strategy for cattle urinary metabolome analysis. *Metabolomics*.

Kaabia, Z., Dervilly-Pinel, G., Popot, M. A., Bailly-Chouriberry, L., Plou, P., Bonnaire, Y., & Le Bizec, B. (2014). Monitoring the endogenous steroid profile disruption in urine and blood upon nandrolone administration: An efficient and innovative strategy to screen for nandrolone abuse in entire male horses. *Drug Test Anal, 6*(4), 376-388.

Kinkead, R. A., Elliott, C. T., Cannizzo, F. T., Biolatti, B., & Mooney, M. H. (2015). Proteomic identification of plasma proteins as markers of growth promoter abuse in cattle. *Anal Bioanal Chem, 407*(15), 4495-4507.

Kouassi Nzoughet, J., Gallart-Ayala, H., Dervilly-Pinel, G., Biancotto, G., & Le Bizec, B. (2015). Original combination of Hydrophilic Interaction (HILIC) and Reverse Phase (RPLC) High Resolution LC-MS for characterizing lipids profile disruption in serum of anabolic implanted bovines. *Metabolomics, 11*(6), 1884-1895.

Kouassi Nzoughet, J. J., Dervilly-Pinel, G., Chéreau, S., Biancotto, G., Monteau, F., Elliott, C. T., & Le Bizec, B. (2015). First insights into serum metabolomics of trenbolone/estradiol implanted bovines; screening model to predict hormone-treated and control animals' status. *Metabolomics*.

Laparre, J., Kaabia, Z., Mooney, M., Buckley, T., Sherry, M., Le Bizec, B., & Dervilly-Pinel, G. (2017). Impact of storage conditions on the urinary metabolomics fingerprint. *Anal Chim Acta, 951*, 99-107.

Luque de Castro, M. D., & Priego-Capote, F. (2017). The analytical process to search for metabolomics biomarkers. *J Pharm Biomed Anal*.

Martin, J. C., Maillot, M., Mazerolles, G., Verdu, A., Lyan Bernard, L., Migné, C., Defoort, C., Canlet, C., Junot, C., Guillou, C., Manach, C., Jabob, D., Jouan-Rimbaud Bouveresse, D., Paris, E., Pujos-Guillot, E., Jourdan, F., Giacomoni, F., Courant, F., Favé, G., Le Gall, G., Chassaigne, H., Tabet, J.-C., Martin, J.-F., Antignac, J.-P.,

Shintu, L., Defernez, M., Philo, M., Alexandre-Gouaubau, M.-C., Amiot-Carlin, M. J., Bossis, M., Triba, M. N., Stojilkovic, N., Banzet, N., Molinié, R., Bott, R., Goulitquer, S., Caldarelli, S., & Rutledge, D. N. (2015). Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. *Metabolomics, 11*, 15.

Milgrom, E., & Baulieu, E. E. (2008). *Hormones, Santé Publique et Environnement* (Vol. Rapport sur la Science et la Technologie, n°28). Les Ulis: Académie des Sciences.

MSI, B. M., Sansone, S., Fan, T., Goodacre, R., Griffin, J., Hardy, N., Kaddurah-Daouk, R., Kristal, B., Lindon, J., Mendes, P., Morrison, N., Nikolau, B., Robertson, D., Sumner, L., Taylor, C., van der Werf, M., van Ommen, B., & Fiehn, O. (2007). The metabolomics standards initiative. *Nat Biotechnol, 25*(8), 846-848.

Naz, S., Vallejo, M., García, A., & Barbas, C. (2014). Method validation strategies involved in non-targeted metabolomics. *Journal of Chromatography A, 1353*, 7.

Nebbia, C., Urbani, A., Carletti, M., Gardini, G., Balbo, A., Bertarelli, D., & Girolami, F. (2011). Novel strategies for tracing the exposure of meat cattle to illegal growth-promoters. *Vet J, 189*(1), 34-42.

Oms-Oliu, G., Odriozola-Serrano, I., & Martín-Belloso, O. (2013). Metabolomics for assessing safety and quality of plant-derived food. *Food Research International, 54*(1), 1172-1183.

Pielaat, A., Barker, G., Hendriksen, P., Hollman, P., Peijnenburg, A., & Ter Kuile, B. (2013). A foresight study on emerging technologies: State of the art of omics technologies and potential applications in food and feed safety. REPORT 1: Review on the state of art of omics technologies in risk assessment related to food and feed safety. . *EFSA supporting publication, EN-495*, 126.

Pinel, G., Weigel, S., Antignac, J. P., Mooney, M. H., Elliott, C., Nielen, M. W. F., & Le Bizec, B. (2010). Targeted and untargeted profiling of biological fluids to screen for anabolic practices in cattle. *TrAC Trends in Analytical Chemistry, 29*(11), 1269-1280.

Pinel, G., Weigel, S., Lommen, A., Chéreau, S., Rambaud, L., Essers, M., Nielen, M., & Le Bizec, B. (2011). Assessment of two complementary LC-HRMS metabolomics strategies for the screening of anabolic steroid treatment in calves. *Anal Chim Acta, 700*(1-2), 10.

Reiter, M., Walf, V. M., Christians, A., Pfaffl, M. W., & Meyer, H. H. (2007). Modification of mRNA expression after treatment with anabolic agents and the usefulness for gene expression-biomarkers. *Anal Chim Acta, 586*(1-2), 73-81.

Riedl, J., Esslinger, S., & Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Anal Chim Acta*.

Riedmaier, I., Tichopad, A., Reiter, M., Pfaffl, M. W., & Meyer, H. H. (2009). Identification of potential gene expression biomarkers for the surveillance of anabolic agents in bovine blood cells *Anal Chim Acta, 638*(1), 106-113.

Rocca-Serra, P., Salek, R. M., Arita, M., Correa, E., Dayalan, S., Gonzalez-Beltran, A., Ebbels, T., Goodacre, R., Hastings, J., Haug, K., Koulman, A., Nikolski, M., Oresic, M., Sansone, S. A., Schober, D., Smith, J., Steinbeck, C., Viant, M. R., & Neumann, S. (2016). Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics, 12*(1), 14.

Wang, X., Wang, S., & Cai, Z. (2013). The latest developments and applications of mass spectrometry in food-safety and quality analysis. *TrAC Trends in Analytical Chemistry, 52*(0), 170-185.

470

471 *Table 1. Details of animal experiments protocols used to assess biomarkers robustness (urine samples)*

Exp.	Animals	Breed	Country	Growth promoters		Treatment
A	18 Calves (12 C / 6 T)	Holstein	FR	β–agonists	Clenbuterol	Oral, 8 µg kg^{-1} bw, 21 days
B	17 Calves (12 C / 5 T)	Holstein	FR	β–agonists	Clenbuterol	Oral, 6 µg kg^{-1} bw, 6 days
C	1 Calf	Montbeliard	FR	β–agonists	Clenbuterol, ractopamine, cimaterol, zilpaterol, mabuterol	Oral, 1 µg kg^{-1} bw each, 5 days
D	1 Heifer	Holstein	FR	β–agonists	Clenbuterol, Ractopamine	Oral, C: 10 µg kg^{-1} bw or R:200 µg kg^{-1} bw or C+R 1 µg kg^{-1} bw and 20 µg kg^{-1} bw, once
E	12 Calves (4 C / 8 T)	Friesian	IT	β–agonists	Clenbuterol	Oral, 8 µg kg^{-1} bw, 21 days
F	24 Calves (12 C / 12 T)	Limousine	NL	Steroids	Estradiol benzoate + Nortestosterone laureate	IM, EBz: 25 mg, NT: 150 mg
G	12 calves (6C /6 T)	Montbeliard	FR	Steroids	Boldenone undecylenate	IM, 2 mg kg^{-1} bw, once
H	17 Calves (2 C / 15T)	Friesian	IT	Corticosteroids	Dexamethasone + Prednisolone Prednisolone	Oral, D: 0.2 mg/day, P: 4 mg/day, 20 days or Oral, 8 mg/day, 20 days

472

473

474

475 *Table 2. Details of bovine animals sampled during the 2013 French national monitoring plan*

Parameters	Categories	Repartition per category (%)
Age	< 6 months	33.2
	6-24 months	24.8
	> 24 months	36.4
	Information not provided	5.60
Gender	Males	50.0
	Females	46.3
	Information not provided	3.70
Production type	Dairy	11.7
	Lactating	38.8
	Suckling calves	1.40
	Calves	15.4
	Information not provided	32.7

476

477

478

479

480

481 *Table 3. Method performance (target compounds = β-agonists, population = calves (i.e. bovine < 6 months),*
482 *matrix = urine, detection time window = 5 days after the end of treatment, threshold: T= $\mu_{control} + 2\sigma$)*

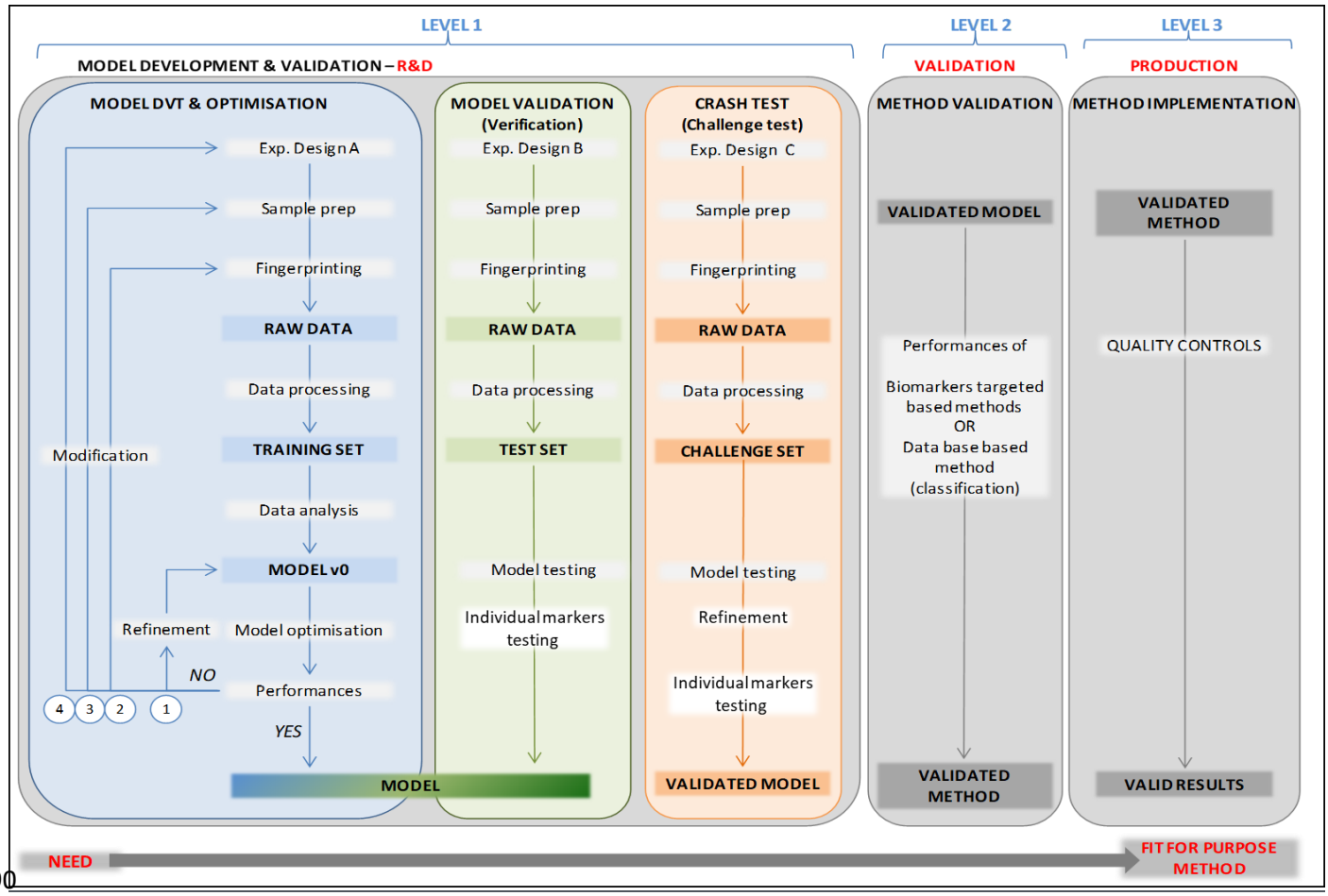
Exp.	Number of samples	False Positive	False negative
A	213	2	3
B	89	6	0
C	10	0	3
D	6	0	2
E	128	3	0
General calves' population	71	14	0
TOTAL	**517**	**25 (4.8%)**	**8 (1.5%)**

483

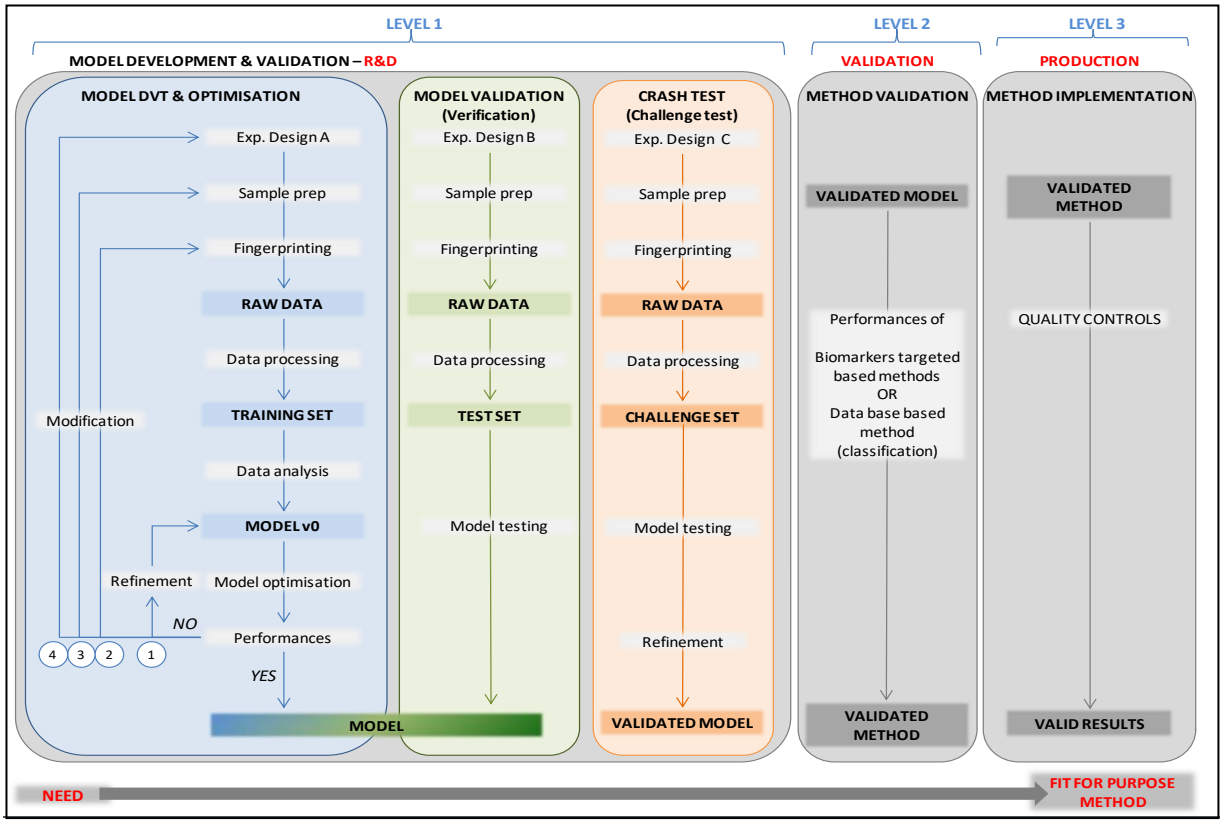
484

485

FIGURE 1


LEVEL 1
LEVEL 2
LEVEL 3
MODEL DEVELOPMENT & VALIDATION – R&D
VALIDATION
PRODUCTION
MODEL DVT & OPTIMISATION
MODEL VALIDATION (Verification)
CRASH TEST (Challenge test)
METHOD VALIDATION
METHOD IMPLEMENTATION
Exp. Design A
Exp. Design B
Exp. Design C
Sample prep
Fingerprinting
RAW DATA
Data processing
TRAINING SET
TEST SET
CHALLENGE SET
Modification
Data analysis
MODEL v0
Refinement
Model optimisation
Performances
NO
YES
4
3
2
1
Model testing
Individual markers testing
Refinement
MODEL
VALIDATED MODEL
VALIDATED METHOD
Performances of
Biomarkers targeted based methods OR Data base based method (classification)
QUALITY CONTROLS
VALID RESULTS
NEED
FIT FOR PURPOSE METHOD

490


LEVEL 1
LEVEL 2
LEVEL 3
MODEL DEVELOPMENT & VALIDATION – R&D
VALIDATION
PRODUCTION
MODEL DVT & OPTIMISATION
MODEL VALIDATION (Verification)
CRASH TEST (Challenge test)
METHOD VALIDATION
METHOD IMPLEMENTATION
Exp. Design A
Exp. Design B
Exp. Design C
Sample prep
Fingerprinting
RAW DATA
Data processing
TRAINING SET
TEST SET
CHALLENGE SET
Modification
Data analysis
MODEL v0
Refinement
Model optimisation
Performances
NO
YES
4
3
2
1
Model testing
Refinement
MODEL
VALIDATED MODEL
VALIDATED METHOD
Performances of
Biomarkers targeted based methods OR Data base based method (classification)
QUALITY CONTROLS
VALID RESULTS
NEED
FIT FOR PURPOSE METHOD

491

492

493
494

FIGURE 2

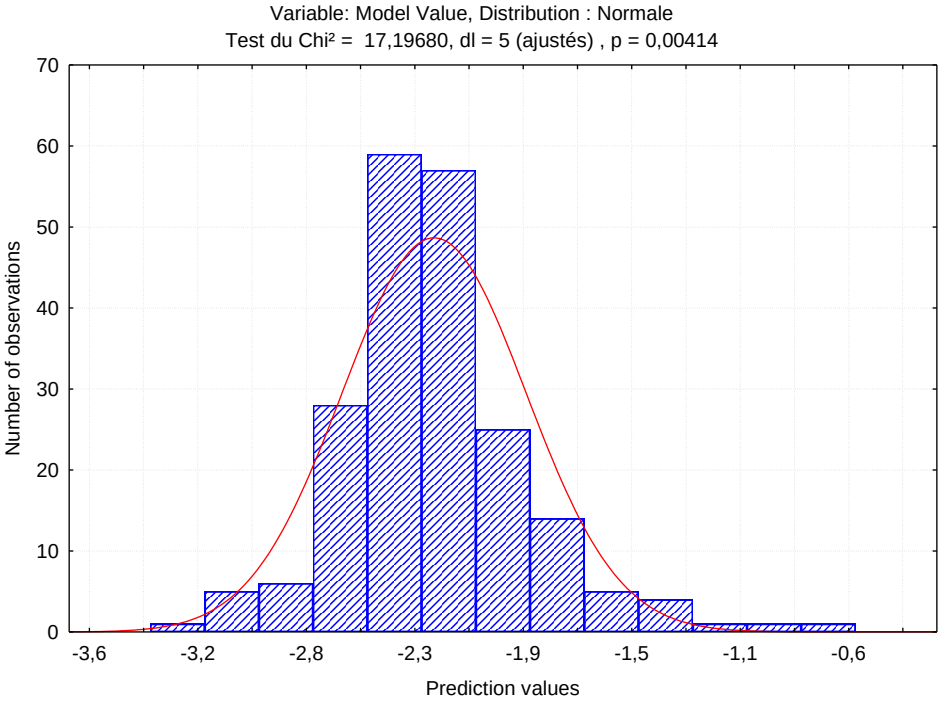

Variable: Model Value, Distribution : Normale
Test du Chi² = 17,19680, dl = 5 (ajustés) , p = 0,00414
Number of observations
70
60
50
40
30
20
10
0
-3,6
-3,2
-2,8
-2,3
-1,9
-1,5
-1,1
-0,6
Prediction values

Figure 3

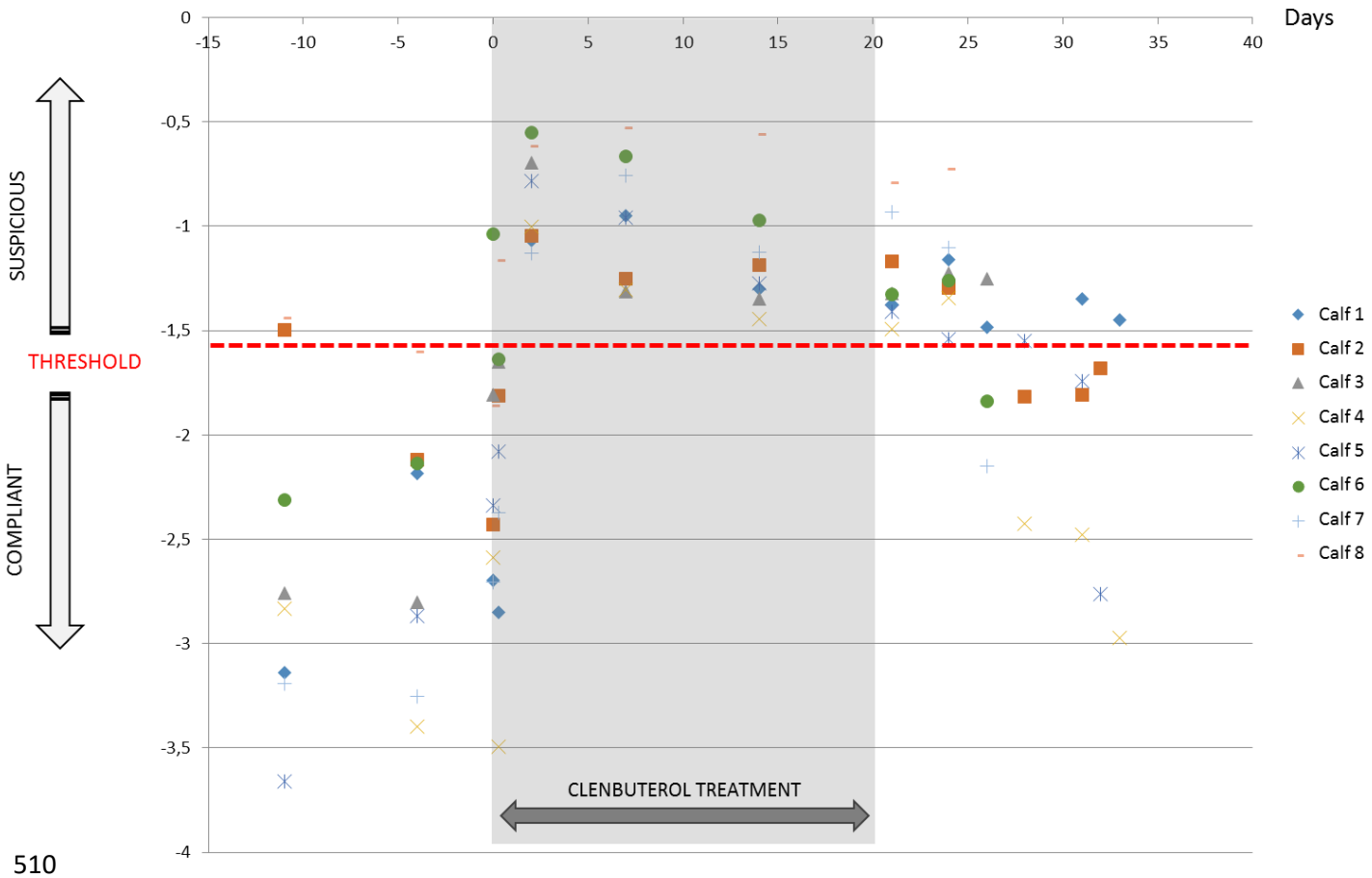

Days
0
-15
-10
-5
0
5
10
15
20
25
30
35
40
-0,5
-1
-1,5
-2
-2,5
-3
-3,5
-4
SUSPICIOUS
THRESHOLD
COMPLIANT
CLENBUTEROL TREATMENT
Calf 1
Calf 2
Calf 3
Calf 4
Calf 5
Calf 6
Calf 7
Calf 8

Figure 4

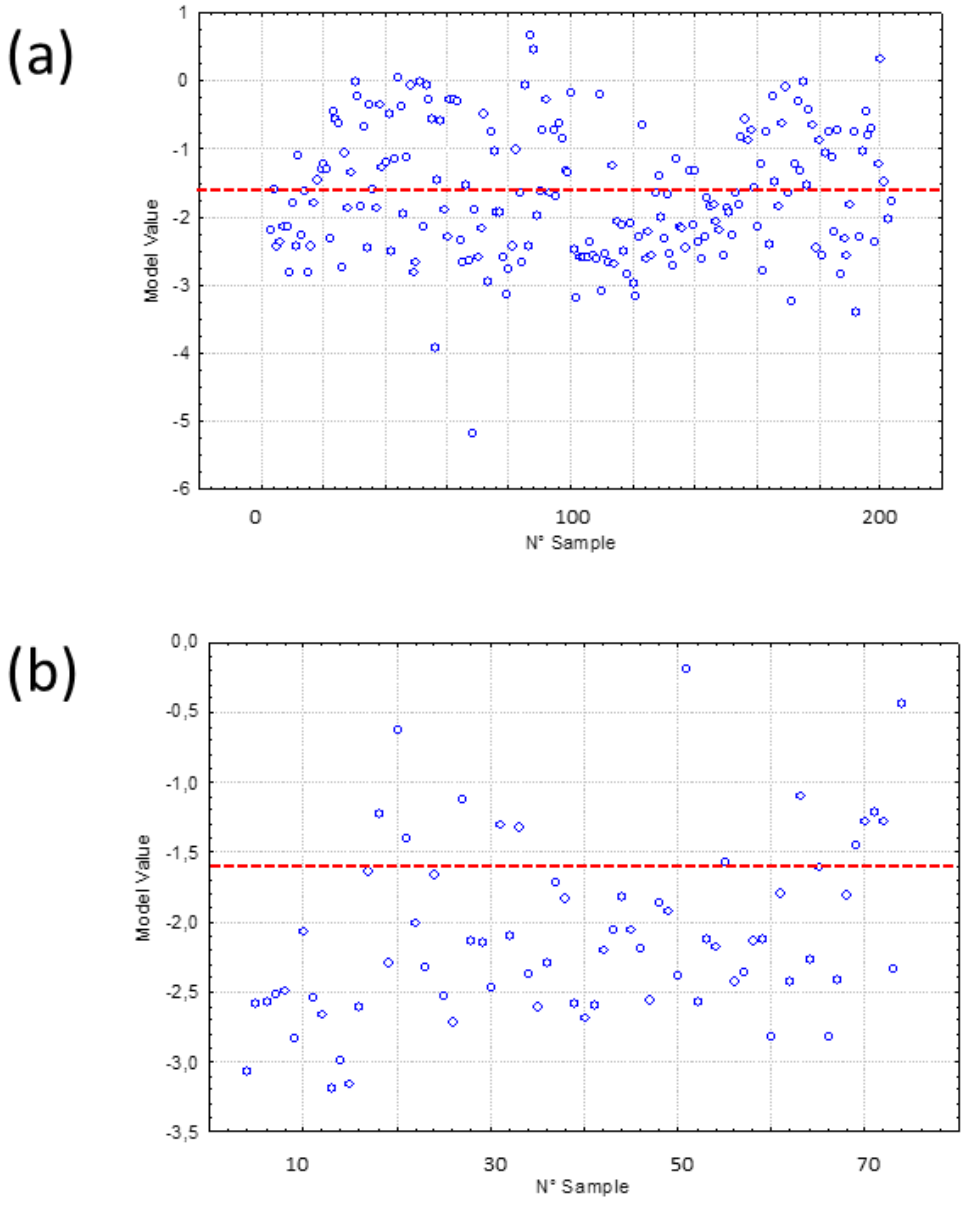


Figure 5

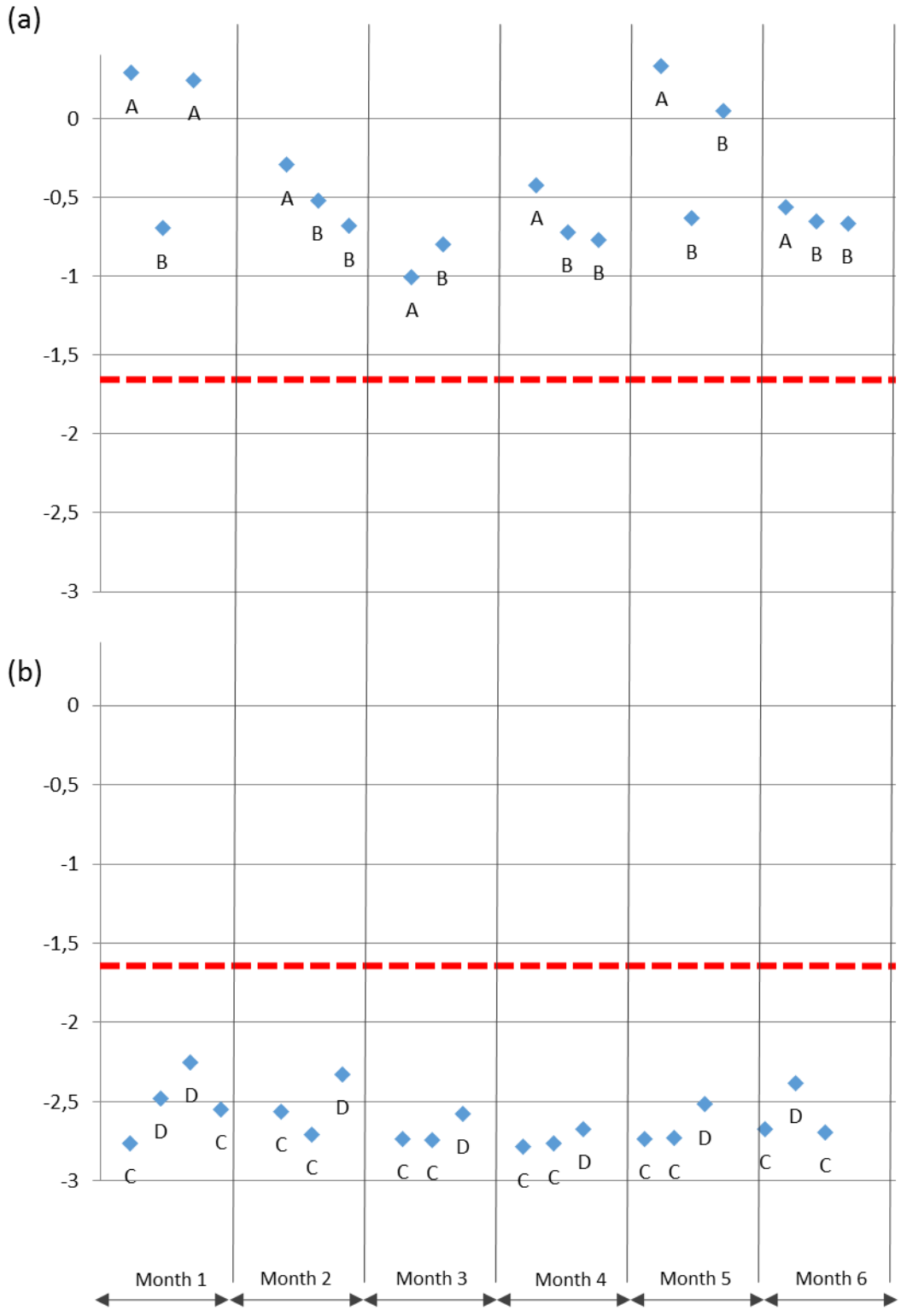

(a)
0
-0,5
-1
-1,5
-2
-2,5
-3
A A B
A B B
A B
A B B
A B
A B B
(b)
0
-0,5
-1
-1,5
-2
-2,5
-3
C D C
C C D
C C D
C C D
C C D
C D C
Month 1
Month 2
Month 3
Month 4
Month 5
Month 6

FIGURE CAPTIONS

Figure 1: Generic scheme for development and validation of untargeted methods

Figure 2: Distribution of prediction values on the model for control samples (n=207)

Figure 3: Prediction of urine samples collected in the frame of Exp. E

Figure 4: Prediction of urine samples collected (a) on general bovine population, (b) on calves

Figure 5: Reproducibility of sample prediction over six months, studied with (a) two non-compliant samples (A and B) and (b) two compliant samples (C and D). Predictions were as follows: A (μ=-0,56; s=0,51), B (μ=-0,66; s=0,24), C (μ=-2,69; s=0,07), D (μ=-2,38; s=0,14).

Supplementary Information

Figure S1: Typical analytical run illustrating the panel of quality assurance points implemented within accredited screening method