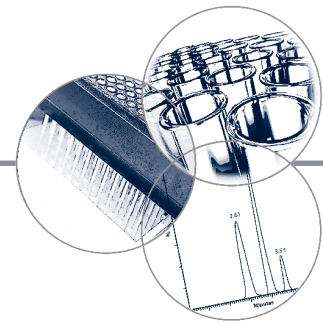
RESEARCH ARTICLE



Incurred sample reanalysis: different evaluation approaches on data obtained for spironolactone and its active metabolite canrenone

Background: The inherent reproducibility of a bioanalytical approach is usually sustained through incurred sample reanalysis (ISR). Questions relating to the number of ISRs, the right moment for performing reanalysis, the way of performing an appropriate statistical refinement of experimental data and actions to be taken in the case of failure are frequently raised. **Results:** Data resulting from ISR following a bioequivalence study for spironolactone formulations are discussed. Reanalysis of samples was carried out twice: immediately after the end of the study and after a period that overcame the long-term stability study achieved during method validation. The Bland–Altman approach was used to assess experimental results. ISR was successful over the short reanalysis period for both compounds. Data produced through reanalysis after the long-term period indicated a systematic positive bias for the metabolite canrenone (although results supported reproducibility). The results obtained for spironolactone were affected by a strong negative systematic bias and failed to support reproducibility. The explanation deals with the continuous conversion of spironolactone to canrenone in plasma samples. However, reproducibility of the method may be sustained by comparing original and repeated differences between concentration values in samples by means of a paired t-test, Wilcoxon sign rank-sum test and linear regression. **Conclusions:** Different statistical approaches for making data comparisons are discussed and may be successfully applied during reanalysis of samples from a bioequivalence study. Results of the evaluations may differ in accordance with the statistical procedure being applied, thus a definitive conclusion requires consideration of all specific experimental circumstances arising during production of the processed data.

The purpose for achieving **incurred sample reanalysis** (ISR) is to reinforce confidence in the reliability and reproducibility of a validated bioanalytical method designed for nonclinical or clinical studies. The topic already has quite a long history, being indissolubly related to bioanalytical method validation and regulation in the respective field [1]. Although regulatory agency guidelines for bioanalytical method validation are already in place [2,3], consistent debates on the ISR subject are extremely factual [4–8]. However, consensus seems to have been reached, at least concerning the following main aspects (only those specifically relating to bioequivalence will be highlighted here) [6]:

- ISR should be performed for all bioequivalence studies;
- ISR should be a part of the process check and should be included in the study report;
- More subjects with fewer samples per subject, with at least one sample close to C_{max} and one near the end of the elimination phase, should be considered;
- Reanalyzed incurred samples should represent between 5 and 10% of the total number of study samples;
- As acceptance criteria specifically dealing with small molecules, two-thirds of repeats agree within ±20% (still to be decided if the agreement is referred to the original value or to the mean of the experimental data pair);
- ISR assessment only is not a sufficient reason to accept/reject a study; however, ISR failure must lead to investigation while the study is kept on hold until investigations are completed and follow-up actions implemented.

ISR may be considered as a particular form of crossvalidation, as it evaluates the results of a bioanalytical method produced during two separate occasions. Basically, the treatment of crossvalidation data involves regression [9–11]. An alternative for evaluation of the agreement between paired datasets is the **Bland–Altman approach** [12–15]. A modified Bland–Altman statistical approach was suggested by Rocci *et al.* [16], based on the assay variability of the bioanalytical method, including both random and **systematic errors**.

Victor Voicu[1], Mihaela Cristina Gheorghe[2], Iulia Daniela Sora[3], Costel Sârbu[4] & Andrei Medvedovici[†2,3]

[1]University of Medicine & Pharmacy 'Carol Davilla', Department of Pharmacology, Toxicology & Clinical Psychopharmacology, #8 Floreasca St, Bucharest, 014461, Romania
[2]University of Bucharest, Faculty of Chemistry, Department of Analytical Chemistry, #90 Panduri Av., Bucharest, 050663, Romania
[3]Bioanalytical Laboratory, SC Labormed Pharma SA, #44B Th. Pallady Blvd, Bucharest, 032266, Romania
[4]Babes-Bolyai University, Faculty of Chemistry & Chemical Engineering, #11 Arany Janos St, Cluj-Napoca, 400028, Romania
[†]Author for correspondence:
Tel.: +40 214 102 279
E-mail: avmedved@yahoo.com

FUTURE SCIENCE part of fsg

Key Terms

Incurred sample reanalysis: Confirmatory action for sustaining the reproducibility of a bioanalytical method, based on a statistical comparison between results obtained on same study samples analyzed during separate occasions.

Bland–Altman approach: Statistical outcome used for evaluation of the agreement between two sets of data representing measurements on paired samples, based on the representation of the difference of each data pair against the average of the two values.

Systematic errors: Bias in measurement leading to measured values being systematically too high or too low with respect to the true value.

Spironolactone: Synthetic 17-lactone drug that is a renal competitive aldosterone antagonist belonging to the class of potassium-sparing diuretics.

Canrenone: Active form and first metabolite of spironolactone.

Linear regression analysis: Evaluation of data from measurements on paired samples, based on the reciprocal plot of data pairs and the statistical analysis of the least square regression parameters characterizing the correlation.

Wilcoxon sign rank-sum test: A distribution-free method for testing the difference between two populations using matched samples. The test is based on the absolute differences of the pairs of observations in the two samples, ranked according to size, with each rank being given the sign of the original difference. The test statistic is the sum of the positive ranks.

Enhancement of the Bland–Altman approach through combination with tolerance intervals in ISR was also proposed [17]. However, only a few applications of ISR for small or large molecules have been published until now [18–24].

The case study presented herein refers to a bioanalytical method validated for the assay of **spironolactone** and its active metabolite **canrenone** for bioequivalence purposes. ISR was successfully performed immediately after study completion on samples from a limited number of volunteers (three from a total of 26). At 9 months after study completion (a period largely overcoming the long-term stability period considered during method validation), the ISR was repeated on samples from all volunteers. Evaluation of data according to the Bland–Altman approach supported the reproducibility of the bioanalytical method toward the metabolite canrenone, but was unsuccessful for spironolactone. In both cases, systematic errors appear (positive for canrenone and negative for spironolactone). Evidently, stability of the target compounds during long-term storage was incriminated. However, the reproducibility of the method may still be sustained through interpretation of the original and repeated differences between concentrations determined for paired samples belonging to a volunteer at the same sampling moment over the two phases of the study. The **linear regression analysis** applied to differences may overcome the systematic errors determined by the transformation of the parent compound to the metabolite on storage and inherently sustained reproducibility. As the decision of bioequivalence between two pharmaceutical formulations is taken through comparison of the concentration profiles for each individual volunteer, the evaluation of the reproducibility may be based on comparison of concentration differences and not of the absolute concentration values. Such an approach circumvents systematic errors produced by different sources (i.e., sample stability problems and equipment variability over time).

In this order, a statistical approach was formulated and the concentration differences were evaluated using the paired t-test, **Wilcoxon sign rank-sum test** and linear regression. The statistical results obtained by applying both tests confirm the hypothesis that the original and repeated differences do not differ significantly. The linear regression indicates some difference in both cases.

Experimental

Reagents

All solvents were HPLC grade from Merck (Darmstadt, Germany). Water for chromatography (resistivity minimum 18.2 MΩ and TOC maximum 30 ppb) was produced within the laboratory by means of a TKA Lab HP 6UV/UF instrument and used during experiments. Spironolactone and canrenone were certified reference substances from the European Pharmacopoeia (Strasbourg, France), batch numbers 1b and 3g, respectively. The internal standard (IS) 7-nitro-5-phenyl-1,3-dihydro-2*H*-1,4-bezodiazepin-2-one/nitrazepam was also a European certified reference substances, batch no. 1b, from the same source.

Apparatus

Experiments were performed with an Agilent 1200 SL series LC–MSD (Agilent Technologies) system consisting of the following modules: degasser (G1322A), binary pump (G1312B), automated injector (G1367C) and the corresponding thermostat (G1330B), column thermostat (G1316B), AP-ESI standard interface (G1948B) and triple quadrupole mass spectrometric detector (G2571A). System control, data acquisition and interpretation were made with the Agilent MassHunter software version B 01.00 incorporating both qualitative and quantitative packages. The system was operationally qualified before and after the bioequivalence study. The vortex system was model Multi Reax from Heidolph (Schwabach, Germany) and the thermostated centrifuge was model Universal 320R from Hettich (Tuttlingen, Germany).

Sample preparation

Plasma sample (0.2 ml) was combined with 0.4 ml of acetonitrile containing the IS at a concentration level of 25 ng/ml. The mixture was vortexed for 5 min at 2000 rpm, and then centrifuged at 7500 × g and 25°C for 5 min. The supernatant was quantitatively collected and diluted with 0.4 ml of water, vortexed at 800 rpm for 10 s and then transferred to the injection vial.

Chromatographic method

A Zorbax Eclipse XDB-C18, 150 mm length, 4.6 mm internal diameter and 3.5 µm particle size (Agilent Technologies, cat. no. 963967–902), fitted with a Phenomenex Guard Cartridge C18, 4 mm × 2 mm (prod. no. AJO-4286) was used. The column was thermostated at 35°C.

Isocratic elution was applied, using a mixture of aqueous 0.1% formic acid and acetonitrile in the volumetric ratio of 30/70 at 0.8 ml/min flow rate. The injected volume was 100 µl. Prepared samples were thermostated in the autosampler during the working sequences at 25°C. As the same mass transitions were used for spironolactone and canrenone, the chromatographic resolution between compounds is absolutely necessary.

MS parameters

The parameters controlling the ESI ion source were as follows: drying gas (N_2) temperature: 350°C; drying gas flow: 13 l/min; pressure of the nebulizing gas: 60 psi; capillary voltage: 2500 V. The fragmentor potential was set at 125 V. Collision-induced dissociation energy (CID) was 30 V for all analytes, using N_2 as the collision gas. The MS/MS detection was carried out in the multiple reactions monitoring (MRM) mode.

Spironolactone does not produce the protonated molecular ion $[M + H]^+$ within the ESI source. Adducts with sodium and potassium are already formed in the source (relative abundances of 90 and 40%, respectively) but could not be used as precursor ions because of the poor precision of the assay. Spironolactone is cleaved to canrenone within the source, consequently producing the corresponding protonated ion. The mass transitions being used for quantitation were *m/z* 341 to 107 for spironolactone and canrenone and *m/z* 282 to 180 for the IS. The mass transitions used for confirmation (qualifier) were: *m/z* 341 to 187 for spironolactone and canrenone and *m/z* 282 to 207 for the IS.

Method validation

Quality attributes of the bioanalytical method are listed in Table 1.

Methodology & pharmacokinetic parameters

In this open-label, randomized, two-period, two-sequence, crossover study, 26 healthy volunteers (male/female) under fasting conditions received one dose of 100 mg of spironolactone of the tested product and one dose of the reference product, in the sequence determined by randomization, with a 10-day wash-out period between consecutive administrations. Only 25 volunteers completed the study. The reference product was Aldactone 100 mg capsules (Roche, Austria). All volunteers signed an informed written consent before initiation of the screening procedure. The protocol of the study was formally accepted by the evaluation department of the Romanian National Drug Agency and received the approval of the Institutional Ethics Committee. Venous blood samples were collected predose (0 h) and at the following post-dose intervals of time: 0.25; 0.5; 0.75; 1.0; 1.25; 1.5; 2; 2.5; 3; 3.5; 4; 4.5; 5; 6; 8; 10; 12; 14; 24; 48; 60 and 72 h.

Evaluation of bioequivalence was made based on primary pharmacokinetic parameters: area under plasma concentration–time plot extrapolated to infinity ($AUC_{infinity}$); area under plasma concentration–time plot until the last quantifiable value (AUC_{last}); and observed maximum plasma concentration for spironolactone and canrenone (C_{max}). The pharmacokinetic parameters were statistically treated by means of ANOVA after log transformation of data for primary pharmacokinetic parameters, with determination of 90% confidence intervals for the intra-individual ratios test/reference products. Acceptance range for concluding bioequivalence was the conventional 90% confidence interval of 80–125% around the geometric mean ratios test/reference of log transformed data for AUC_{last}, $AUC_{infinity}$ and C_{max}. Pharmacokinetic parameters were determined by means of the Kinetica™ software (version 4.4.1.) from Thermo Electron Corporation, USA, and are given in Table 2.

Results & discussions

Incurred sample reanalysis has been achieved over two stages. The first reanalysis was made 1 week after the bioanalytical study completion and is further referred to as the 'short-term study'. Samples being reanalyzed belong to volunteers 1 (phase 1, sampling moments from 1 to 6 h), 10 (phase 1, sampling moments from 0.5 to 5 h and phase 2, sampling moments from 0.25 to 6 h) and 26 (both phases, sampling moments from 0.5 to 8 h), representing a total of 65 (5.9% from all study samples). Original results (further denoted by O) were all placed above the method's LLOQ for both target analytes.

The second reanalysis was made 9 months after the study completion. This second stage aimed to emphasize the method's reproducibility over a longer period after its initial application, also taking into consideration the new issued regulatory guidances about ISR, extensively discussed in the bioanalytical community. The moment of reanalysis was placed outside the 3-month stability interval determined during

Table 1. Quality characteristics of the analytical method, as resulting from the validation procedure and the study completion.

Parameters/quality attributes	Spironolactone	Canrenone
Linearity		
Concentration levels (ng/ml)	4/10/25/50/75/100/125/150	4/25/50/100/150/200/300/400
Independent replicates per concentration level	6	6
Back-interpolated concentration values: mean% bias	-13.5–11.5	-12.9–8.0
Back-interpolated concentration values: mean precision (%RSD)	0.98–3.5	0.97–2.9
Response function	Linear, weighted 1/x	Linear, weighted $1/x^2$
Slope ($B \pm s_B$)	0.0058 ± 0.0001	0.0125 ± 0.0004
Intercept ($A \pm s_A$)	0.0032 ± 0.0041	-0.0074 ± 0.0045
Correlation coefficient	0.9988	0.9969
LLOQ ($5 \times s_A/B$)/method's LLOQ (ng/ml)	3.5/4	1.8/4
Concentration (ng/ml) for signal to noise (S:N) ratio 10:1	1.3	1.0
Selectivity		
% residual area of the blanks compared with area at method's LLOQ (n = 6)	min = 4.3	min = 1.6
	max = 10.3	max = 4.9
Matrix effect (n = 6, including hemolyzed and hyperlipemic blanks)	0.94 (0.84–1.05) %RSD = 6.1	0.85 (0.81–0.88) %RSD = 3.5
IS normalized matrix effect	0.86 (0.77–0.91) %RSD = 5.64	0.78 (0.74–0.83) %RSD = 5.1
Recovery		
Residual plasma matrix effect on ionization (post-spiked plasma vs water-spiked samples; %)	105.8	81.5
Adsorption on precipitated proteins (spiked plasma vs post-spiked plasma samples; %)	88.8	105.2
Precision		
Concentration levels (ng/ml)	12/20/60/120	12/60/180/360
Intra-day (n = 10, independently prepared samples) %RSD	1.7–10.1	1.1–3.0
Intra-day (n = 10, independently prepared samples) % bias	-12.1–6.5	-8.6–12.9
Inter-day (n = 6) %RSD	5.3–10.0	3.3–5.1
Inter-day (n = 6) % bias	-8.0–-1.1	-8.8–8.4
Stability		
Freeze–thaw (n = five cycles; three concentration levels) %RSD % bias	 6.3–8.6 -14.8–8.7	 6.0–6.5 -6.0–13.8
Long-term stability: duration: 3 months; samplings = 3 (monthly); conditions: -40°C %RSD % bias	 6.2–9.9 -14.7–8.8	 3.2–6.8 -13.3–7.3
Short-term stability: duration: 24 h; samplings = 4; conditions: 25°C %RSD % bias	 6.7–9.9 -15.0–8.7	 3.5–5.9 -10.0–7.3
Post-preparative stability: duration: 24 h; samplings = 4; conditions: 25°C %RSD % bias	 5.1–10.7 -14.3–8.8	 4.1–5.7 -7.0–11.1
Dilution integrity		
Concentration level (ng/ml) Independent replicates per dilution ratio (1/8, 1/4 and 1/2) Recovery (%) %RSD (n = 18)	200 6 95.2–113.5 8.8	400 6 96.3–106.9 5.9

†Peak area ratio (analyte to internal standard) divided by the calculated concentration value for incurred samples (response variation per unit of concentration).
IS: Internal standard; QC: Quality control; RSD: Relative standard deviation.

Table 1. Quality characteristics of the analytical method, as resulting from the validation procedure and the study completion (cont.).

Parameters/quality attributes	Spironolactone	Canrenone
Bioanalytical study		
No. of analytical sequences	25	
Format of an analytical sequence	Calibration (eight concentration levels as per linearity)/QC set (four concentration levels as per precision)/incurred samples (from one volunteer, increasing order of the sampling time, alternating phases)/QC set	
%RSD for IS peak area (calibrations + QC sets + incurred samples)	8.9 (n = 1554, no trend for variation)	
Slopes of the calibrations (mean value; %RSD)	0.00578 (8.8)	0.01256 (7.3)
Calibrators failing to back-interpolate in the regressions (from 200)	None	None
QC samples failing to interpolate in the regression (from 200)	3 (1 × QC1, 2 × QC2)	None
%RSD for LLOQ samples (calibration level 1; n = 25)	14.2	8.1
Mean % from LLOQ mean peak area in pre-dose collected samples	14.2 (phase 1)/3.1 (phase 2)	9.6 (phase 1)/3.5 (phase 2)
Absolute retention in the chromatograms of incurred samples (mean/%RSD/n)	4.921 min/1.6 /795	5.911 min/1.9/980
IS absolute retention in the chromatograms of incurred samples (mean/%RSD/n)	3.550 min/0.9/980	
Confirmation factor (quantifier/qualifier ratio; mean/%RSD/n)	17.9/8.0/795	16.8/6.5/980
IS confirmation factor (quantifier/qualifier ratio; mean/%RSD/n)	89.2/0.7/980	
Normalized response† in incurred samples (mean value/%RSD/n)	0.00585/10.9/795	0.01217/10.8/980

†Peak area ratio (analyte to internal standard) divided by the calculated concentration value for incurred samples (response variation per unit of concentration).
IS: Internal standard; QC: Quality control; RSD: Relative standard deviation.

method validation for long-term stability of samples stored at -40°C. This time, sampling was made according to the general accepted rules for ISR, meaning that sampling times 1.5 and 6 h from all volunteers and from both phases were reanalyzed (a total of 100 representing 9.1% from study samples). This step will be further addressed in the text as long-term reanalysis.

Interpretation of the experimental data is presented in Table 3. The Bland–Altman approach was used based on two statistical interpretations. The first refers to evaluation of logarithmic (base 10) transformed data (original [O] and repeated [R]). Computational steps are as follows [16,25]: calculation of the log difference (log O - log R); calculation of the mean (D) and standard deviation (s_D) of the differences between logarithmic data; calculation of the mean ratio (MR) as 10^D; calculation of the 95% confidence interval of the mean ratio, through applying the relationship:

$$CLs = 10^{D \pm 2*SD/\sqrt{n}}$$

where, n is the number of paired data; calculation of the acceptance limits:

$$ALs = 10^{D \pm SD}$$

The second interpretation [17,26] deals with calculation of the neat difference (R - O) between paired determinations, followed by normalization of the difference to the average of the paired results. The mean of the normalized differences and the corresponding standard deviation are used for the calculation of the confidence limits as well as the tolerance limits. The tolerance limits (TLs) are expressed as $D \pm k*s_D$, where k is the tolerance factor for a normal distribution and a proportion, p = 66.7%.

Conclusions can be drawn from analyzing data in Table 3 based on log transformed values:

- The reanalysis confirms precision and accuracy of the method towards spironolactone (the short-term approach), as both confidence and acceptance intervals are within 0.83 and 1.2. The confidence interval does not include 1, suggesting a systematic bias of data; indeed, 55 repeated values from 65 are higher than original ones; the bias is positive;
- For the metabolite canrenone, data interpretation leads to a positive conclusion, both confidence and acceptance intervals being included in the 0.83 to 1.2 range. The confidence interval includes 1, sustaining that no significant bias is produced through reanalysis; this was confirmed by the fact that 31 repeated values from 65 are higher than original ones (positive and negative biases are almost equally distributed);

Table 2. Statistic interpretation of the principal pharmacokinetic parameters corresponding to spironolactone and canrenone.

Pharmacokinetic parameter	Reference	Tested	90% confidence limits	%CV within subject
Spironolactone				
C_{max} (ng/ml)	50.65	56.31	0.9596–1.2492	27.20
Standard deviation for C_{max}	27.82	36.34		
AUC_{last} (ng/ml/h)	158.38	154.41	0.9065–1.1671	26.07
Standard deviation for AUC_{last}	88.09	71.63		
$AUC_{infinity}$(ng/ml/h)	184.22	181.22	0.9394–1.1326	19.30
Standard deviation for $AUC_{infinity}$	93.66	75.32		
Canrenone				
C_{max} (ng/ml)	74.91	86.31	1.0463–1.2285	16.56
Standard deviation for C_{max}	32.29	36.00		
AUC_{last} (ng/ml/h)	1413.98	1575.75	1.0246–1.1837	14.89
Standard deviation for AUC_{last}	578.87	642.55		
$AUC_{infinity}$ (ng/ml/h)	1707.70	1856.43	1.0021–1.1568	14.81
Standard deviation for $AUC_{infinity}$	714.94	764.41		

- Reanalysis for spironolactone over the long-term period fails; both confidence and acceptance limits are placed outside the 0.83 to 1.2 range; a significant negative bias could be observed as long as 79 repeated values are lower than the original ones;
- Results obtained on reanalyzing canrenone in samples stored over the long-term period produced a confidence interval within the 0.83 to 1.2 range. The confidence interval does not include 1, a systematic bias being indicated; indeed, 73 paired data from 100 indicate that repeated values are systematically higher than original ones; the acceptance interval is not contained within 0.83 and 1.2 (the inferior acceptance limit is 0.76); 21% from data pairs are placed outside the 0.83 to 1.2 range. It is to assume that despite the non-inclusion of the acceptance interval in the allowed range, the results of the test are positive.

Data in Table 3, resulting from analysis of normalized differences, allow the following considerations:

- For spironolactone over the short-term study, the mean (positive, different from 0) and the confidence interval (not including 0) indicate a systematic positive error. Confidence limits are included in the -0.2 to 0.2 range, sustaining the accuracy of the approach; the upper limit of the tolerance interval is placed above the 0.2 threshold; the vertical spread of error in the data appear more frequently at lower mean values (graph not shown); there are a total of 12 points outside the tolerance interval, four above and eight below, which consistently meet the normal distribution of the outliers; ten of the points fall outside the -0.2/0.2 limits specified by the 4/6/20 rule. If the positive systematic bias was nonexistent (normalized differences are corrected with the mean), random errors would push only three pairs outside the ±20% limit [Medvedovici A *et al.* Unpublished Data]; the coefficient of variation (CV = S_D/ √2) is 7.5%, lower than the commonly accepted threshold in bioanalysis.
- For canrenone over the short-term study, the mean (slightly negative, but very close to 0) and the confidence interval (including 0) indicate the lack of any systematic error. Both confidence and tolerance limits are included in the -0.2 to 0.2 range; the vertical spread of error in the data is independent with respect to the mean values (graph not shown); there are a total of 15 points outside the tolerance interval, eight above and seven below; only three points fall outside the -0.2/0.2 limits; the coefficient of variation is 5.8%.
- For spironolactone over the long-term study, the mean is situated far below 0 (even below the -0.2 threshold) and the confidence interval does not include 0, both indicating a strong negative systematic error. The confidence interval is practically outside the -0.2 to 0.2 range, proving a serious lack of accuracy; tolerance limits largely include the -0.2 to 0.2 range; the

Table 3. Results from processing incurred sample reanalysis data (short- and long-term reanalysis) by means of the Bland–Altman approach.

Study	Range of the geometric means	Mean difference of log (10 base) values	Standard deviation of the log difference	Mean ratio	Confidence limits of ratios	Acceptance limits of ratios	Accepted range	n	No. of negative biases (log R> log O)	No. of data pairs out of accepted range (%)	No. of data pairs out of acceptance limits (%)
Spironolactone short-term reanalysis	Min: 3.5 Max: 44.3	-0.0478	0.0465	0.8958	Min: 0.8723 Max: 0.9198	Min: 0.8043 Max: 0.9969	Min: 0.83 Max: 1.20	65	55	16 (24.6)	23 (35.4)
Spironolactone long-term reanalysis	Min: 2.2 Max: 136.6	0.1110	0.2112	1.2913	Min: 1.1716 Max: 1.4232	Min: 0.7940 Max: 2.0999	Min: 0.83 Max: 1.20	100	21	63 (63.0)	16 (16.0)
Canrenone short-term reanalysis	Min: 1.4 Max: 86.1	0.0037	0.0359	1.0086	Min: 0.9882 Max: 1.0295	Min: 0.9287 Max: 1.0954	Min: 0.83 Max: 1.20	65	31	3 (4.6)	18 (27.7)
Canrenone long-term reanalysis	Min: 6.5 Max: 156.2	-0.0315	0.0829	0.9299	Min: 0.8951 Max: 0.9662	Min: 0.7683 Max: 1.1256	Min: 0.83 Max: 1.20	100	73	21 (21.0)	13 (13.0)
	Range of the means	**Mean difference normalized to the average**	**Standard deviation of the normalized mean difference**	**No. of data pairs out of the interval ±20% compared with the original value (%)**	**Confidence limits of the normalized difference**	**Tolerance limits of the normalized difference**	**Accepted range**	**n**	**No. of negative biases (R<O)**	**No. of data pairs out of accepted range (%)**	**No. of data pairs out of tolerance limits (%)**
Spironolactone short-term reanalysis	Min: 3.5 Max: 44.4	0.1097	0.1063	15 (23.1)	Min: 0.0838 Max: 0.1355	Min: -0.0118 Max: 0.2311	Min: -0.2 Max: 0.2	65	10	10 (15.4)	12 (18.5)
Spironolactone long-term reanalysis	Min: 2.2 Max: 134.6	-0.2453	0.4176	56 (56.0)	Min: -0.3272 Max: -0.1634	Min: -0.7055 Max: 0.2149	Min: -0.2 Max: 0.2	100	79	61 (61.0)	17 (17.0)
Canrenone short-term reanalysis	Min: 1.4 Max: 86.2	-0.0085	0.0824	2 (3.1)	Min: -0.0286 Max: 0.0115	Min: -0.1027 Max: 0.0856	Min: -0.2 Max: 0.2	65	34	3 (4.6)	15 (23.1)
Canrenone long-term reanalysis	Min: 6.6 Max: 157.5	0.0715	0.1818	21 (21.0)	Min: 0.0359 Max: 0.1071	Min: -0.1288 Max: 0.2719	Min: -0.2 Max: 0.2	100	27	18 (18.0)	12 (12.0)

Statistics are made on log transformed data (rows 1–4) based on data in [16,25] *and the differences normalized to the mean of the paired determinations (rows 5–8) as described in* [17,26].
O: Original; R: Repeat.

vertical spread of errors in the data evidently affects lower values (graph not shown); while 17 points are placed outside the tolerance interval, 61 values are lower than -0.2 (54) or higher than 0.2 (7). Correction of the normalized differences with the mean to compensate the systematic negative bias produces 49 outliers from the -0.2 to 0.2 interval, suggesting that random errors are also strongly affecting the results; the coefficient of variation is 29.5%, largely overcoming the accepted 15% limit.

- For canrenone over the long-term study, the mean is slightly positive and the confidence limits do not include 0, indicating a positive systematic bias. Confidence interval is included in the acceptance range, while the tolerance interval exceed the upper limit of 0.2; 12 data points (seven below and five above) are placed outside the tolerance interval; 18 values are outside the -0.2 to 0.2 range, 15 data points being placed above the upper limit. Through correction the data points with the observed bias, still 11 outliers from the accepted range are observed, indicating that the random errors effect is not negligible; the percentage coefficient of variation calculated value is 12.9%, quite close to the limit.

As expected, the Bland–Altman approach applied on both log transformed and absolute differences between paired datasets provided similar results. Obviously, ISR applied for canrenone over the short-term study fulfils acceptance of all criteria based on both data interpretation modes: no systematic bias, confidence, acceptance and tolerance limits included in the imposed ranges. Similarly, the two data processing modes lead to the same negative conclusion in case of ISR for spironolactone over the long-term study: strong negative systematic bias, confidence, acceptance and tolerance limits non-fitting to the imposed ranges. ISR carried out for spironolactone over the short-term period fulfilled criteria on interpretation of the log transformed data (confidence and acceptance limits within the 0.83 to 1.2 interval), although a slight positive bias was evidenced. The approach based on mean normalized differences between paired data revealed the same positive bias, inclusion of the confidence limits and the non-inclusion of the tolerance interval within the -0.2 to 0.2 range. The number of outliers agrees with the 4/6/20 rule (only 10% of the results instead of the 33.3% are placed outside the -0.2 to 0.2 interval). For ISR on canrenone over the long-term period, both methodologies indicate an evident positive systematic bias. Confidence intervals are included within agreed ranges, while acceptance and tolerance limits do not match those ranges. Of the reanalyzed samples, 18% are outliers. If the differences between the paired data are compared with the original values and not with the means (according to the opinion of the European Bioanalysis Forum), the 4/6/20 rule is also fulfilled for ISR studies carried out for spironolactone over the short-term period and for canrenone over both short- and long-term periods.

The slight positive bias observed for spironolactone at the short-term interval may be tentatively explained through the modification of the ionization yield produced by the accumulation of the residual matrix in the MS source after completion of the study analysis (it is noteworthy that no source-cleaning operations have been made over completion of the study). From data in Table 1, spironolactone seems to be more sensitive compared with canrenone to matrix effects and the residuals collected in the ion source after injection of 1600 samples may be an explanation of the systematic positive error.

The fact that ISR for spironolactone fails over the long-term may be explained by conversion to canrenone even on storage conditions at -40°C. Instability of spironolactone in rat plasma samples has been reported in the literature already [27]. This produces a systematic strong negative bias during ISR assessment. Errors are mostly affecting the low concentration values. This may be explained if a pseudo first-order kinetic controls the conversion of the parent compound to the metabolite. Although the instrumental limit of quantitation seems to be lower than the method's limit (see Table 1), reduction of the concentration values through conversion to the metabolite add an evident degree of randomness to final results (elimination of the bias by correction of the normalized difference through substraction of the mean still produces 49% outliers). Production of canrenone on storage generates the consistent systematic positive bias observed during evaluation of ISR during the long-term study. However, plasma levels of spironolactone are consistently lower than those of the related active metabolite. A partial conversion of spironolactone is not enough to push canrenone levels outside the accepted range, consequently only 21% of the samples produced reanalysis differences compared with original levels higher than 0.2 (if comparison is made with respect to the mean, the outliers represent 18%).

However, it may be interesting to establish if an 'absolute' accuracy of a bioanalytical method should be considered as mandatory after performing ISR. The verdict of bioequivalence between two pharmaceutical formulations is not taken based on absolute plasma concentration values, but on differences appearing between concentration–time profiles from each of the volunteers participating in the clinical trial. Consequently, a decision about the bioequivalence between two pharmaceutical formulations may be correctly taken when proving the accuracy of the method toward the differences between plasma concentrations values (considering the same sampling moments over the two phases, one phase corresponding to the reference product and the other to the tested one). In such conditions, systematic biases are evidently loosing importance on evaluation and only random errors should be emphasized.

We considered that data produced on reanalysis of study samples over the long-term interval may represent an experimental basis for a statistical evaluation alternative that is not considering the systematic errors produced through conversion of spironolactone to canrenone on storage. Consequently, two different datasets have been produced, through making the difference between plasma concentration values in samples taken at the same sampling time from clinical phases resulting after administration of the tested and reference medications. At 1.5 and 6 h after both administration phases, samples considered for long-term ISR were taken from each volunteer. Through considering the randomization schema used during the clinical trial, it is possible to generate data pairs (n = 50) through substracting the concentration value found in the sample after reference product administration from the concentration value found in the sample after administration of the tested product (T - Ref.). The original set of differences (O) is calculated from data on study completion, the repeated set of differences (R) is calculated from data from reanalysis of these samples after the long-term period.

As the Bland–Altman approach applied to differences between concentrations over the two clinical phases failed for both spironolactone and canrenone, alternative statistical evaluation modes [28–30] are further discussed.

Summary statistics (arithmetic mean and median, minimum, maximum, range, standard error, and also skewness and kurtosis) for the all differences between T and Ref. (O and R) are shown in Table 4 and Figure 1.

Table 4. Statistic results determined on datasets representing the differences between concentrations determined in samples withdrawn after administration of the tested and reference products, at precise sampling moments.

Analyte	Differences (T - Ref.) set	n	Mean	Confidence interval (α = 0.05)	Median	Min.	Max.	Range	Standard error	Skewness	Kurtosis
Spironolactone	Original	50	1.411	-3.675–6.497	-1.598	-43.228	77.841	121.069	2.531	2.352	9.051
	Repeat	50	3.394	-1.130–7.917	-0.438	-21.482	75.658	97.140	2.251	2.669	9.196
Canrenone	Original	50	8.084	2.194–13.974	3.372	-32.913	67.666	100.579	2.931	1.000	1.509
	Repeat	50	10.030	4.958–15.102	7.982	-20.990	55.651	76.641	2.524	0.685	0.202

Ref.: Reference; T: Tested.

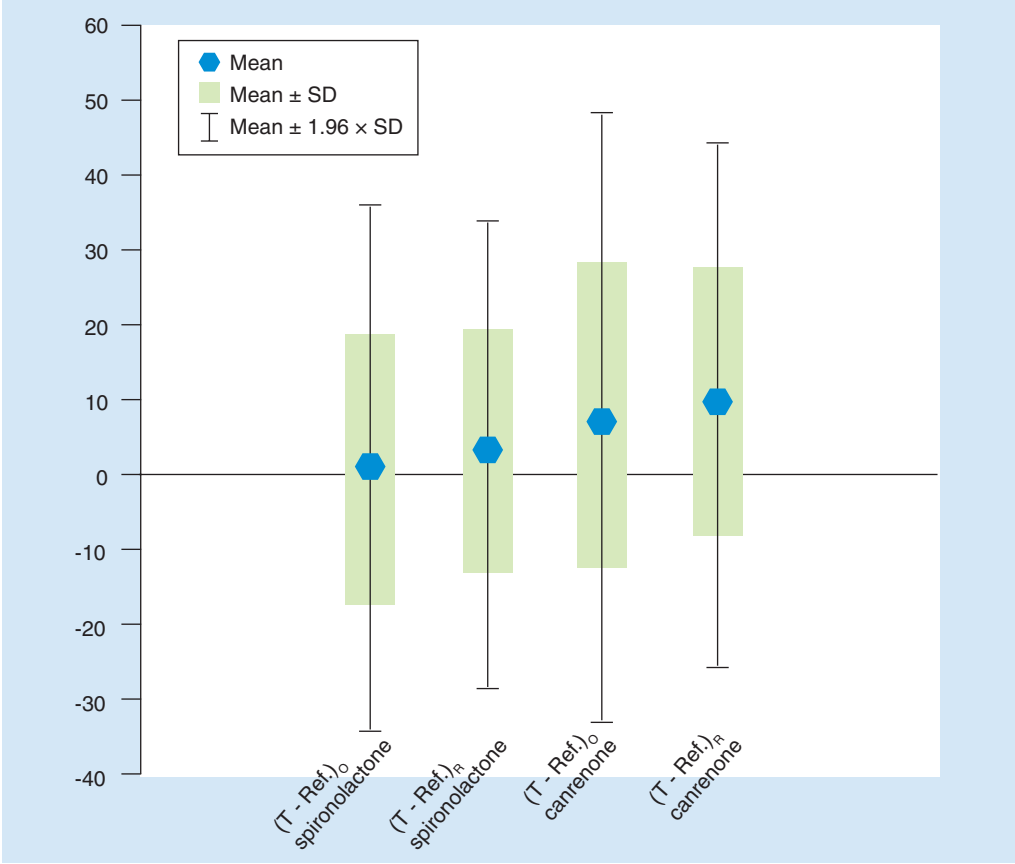


Figure 1. Box and whisker plot of the results obtained for original and repeated differences.
Ref.: Reference; SD: Standard deviation; T: Tested.

Note that the patterns in Figure 1 are in good agreement with the results in Table 4 and suggest that there is a high similarity between the original and repeated differences between test and reference for both compounds concerning mainly the spread of results (standard error, skewness and kurtosis). Taking into account the experimental design and the structure of data, it is evident that the **paired t-test** and nonparametric tests appear to be the most suited to prove the similarities or differences between the results obtained in this study. The paired t-test is based on the differences between the observations of the matched pairs. The statistic test is given by the mean of the differences and their standard deviation. If data are naturally paired, the paired t-test appears to be much more powerful than the two-sample test and should be applied. The statistical results obtained applying the paired t-test (Table 5) indicate that the individual differences are quite similar, that is, difference between the original and repeated (T - Ref.) differences are not significant ($p > 0.35406$ for spironolactone and $p > 0.28402$ in the case of canrenone).

Similar to the paired t-test, the Wilcoxon sign rank-sum test was applied to the analysis of paired/matched original and repeated differences between test and reference measurements. The Wilcoxon sign rank-sum test is a nonparametric counterpart of the paired t-test. However, unlike the parametric tests, the Wilcoxon test does not assume that considered data is normally distributed and is less affected by extreme values. The differences are first ranked without regard to the sign. The smallest difference is given rank 1 and ties are assigned average ranks. To each rank, sign of the difference is assigned. The positive and the negative ranks are summed up and the absolute value of the smaller of the rank (T) is then compared with the standard critical values for testing the hypothesis. Since in our case the number of measurements is large enough, the statistic Z, which is supposed to follow a standard normal distribution, can be used. The results are very similar with those obtained applying the paired t-test and thus strongly support the same conclusion: the difference between the

Key Term

Paired t-test: A Student's t-test for the equality of the means of two populations, when the observations arise as paired samples. The test is based on the differences between the observations of the matched pairs.

Table 5. Results obtained through applying the paired t-test.

Analyte	Data	Mean	SD	n	Diff.	SD of Diff.	t	df	p-value
Spironolactone	$(T - Ref.)_O$	1.411	17.896	50	-1.983	14.984	-0.936	49	0.35406
	$(T - Ref.)_R$	3.394	15.916						
Canrenone	$(T - Ref.)_O$	8.084	20.726	50	-1.946	12.704	-1.083	49	0.28402
	$(T - Ref.)_R$	10.030	17.848						

df: Degrees of freedom; Diff: Difference; O: Original; R: Repeat; Ref.: Reference; SD: Standard deviation; t: t-test value; T: Tested.

original and repeated differences are not significant ($p > 0.47800$ for spironolactone and $p > 0.35659$ in the case of canrenone).

Finally, we applied the linear regression to compare the difference among (T - Ref.) differences (see Figure 2A & B). When plotting the results from one set against those from the other, a straight line is expected with a slope not differing significantly from unity and an intercept not differing significantly from zero. A slope significantly different from unity indicates a proportional difference and an intercept different from zero indicates a constant difference (bias). As the conversion of spironolactone to canrenone during sample storage has been previously appointed as the major reason for the total or partial failure of the Bland–Altman approach carried out on absolute concentration values, it is expected that for the reciprocal plots of the concentration differences recorded during the repeat versus original events for both analytes, proportional errors should be observed; materialized in slopes significantly differing from unit as well as intercepts significantly equal to zero, the systematic bias being compensated.

To check the statistical significance of the linear regression parameters' closeness to 1 and 0, a t-test for the slope and the intercept was applied. Considering the original (T - Ref.) differences as independent variables and repeat (T - Ref.) differences as analytical results for each compound, the slope is significantly different from 1 according to the p-value in Table 6, verifying the above mentioned assumption.

The intercept found for the linear regression belonging to spironolactone is not significantly different from zero, confirming the compensation of the systematic negative bias through substracting between the absolute concentration values. However, this is not the case for canrenone, as the intercept statistically differs from zero. This can be assigned to the major influences of outlier experimental data on the linear regression treatment. For strengthening these assumptions and for identification of outliers, it is important to check if random errors affecting (T - Ref.) differences, approximate to a normal distribution. Probability plots (shown in Figure 2A1 & B1) were used to observe the normal distribution of the experimental errors. The distances between the (T - Ref.) experimental differences and their averaged value were rationed by the standard deviation (s_D):

$$z_i = [(T - Ref.)_i - \overline{(T - Ref.)}]/s_D$$

These values were arranged in increasing order and compared with a set of 50 random values known to be normally distributed with a mean of zero and a standard deviation of 1 (for a detailed computation, see [17]). Analyzing plots from Figure 2A1 & 2B1, it can be advanced that the random error in (T - Ref.) differences obtained for canrenone during the original and repeat events have distributions reasonably close to the normal one (the slopes of the linear regressions in Figure 2B1 are 0.9874 and 1.0014, respectively, and correlation coefficients are

Table 6. Statistical parameters to evaluate linear relation between original and repeated differences.

$(T - Ref.)_O = f(T - Ref.)_R$		Value	SE	t ($\nu = 48$)	p-value
Spironolactone	Intercept	2.625	1.8029	1.4557	0.151964
	Slope	0.545	0.1014	5.3717	0.000002
Canrenone	Intercept	4.509	1.6698	2.7005	0.009536
	Slope	0.683	0.0757	9.0194	0.000000

O: Original; R: Repeat; Ref.: Reference; SE: Standard error; T: Tested.

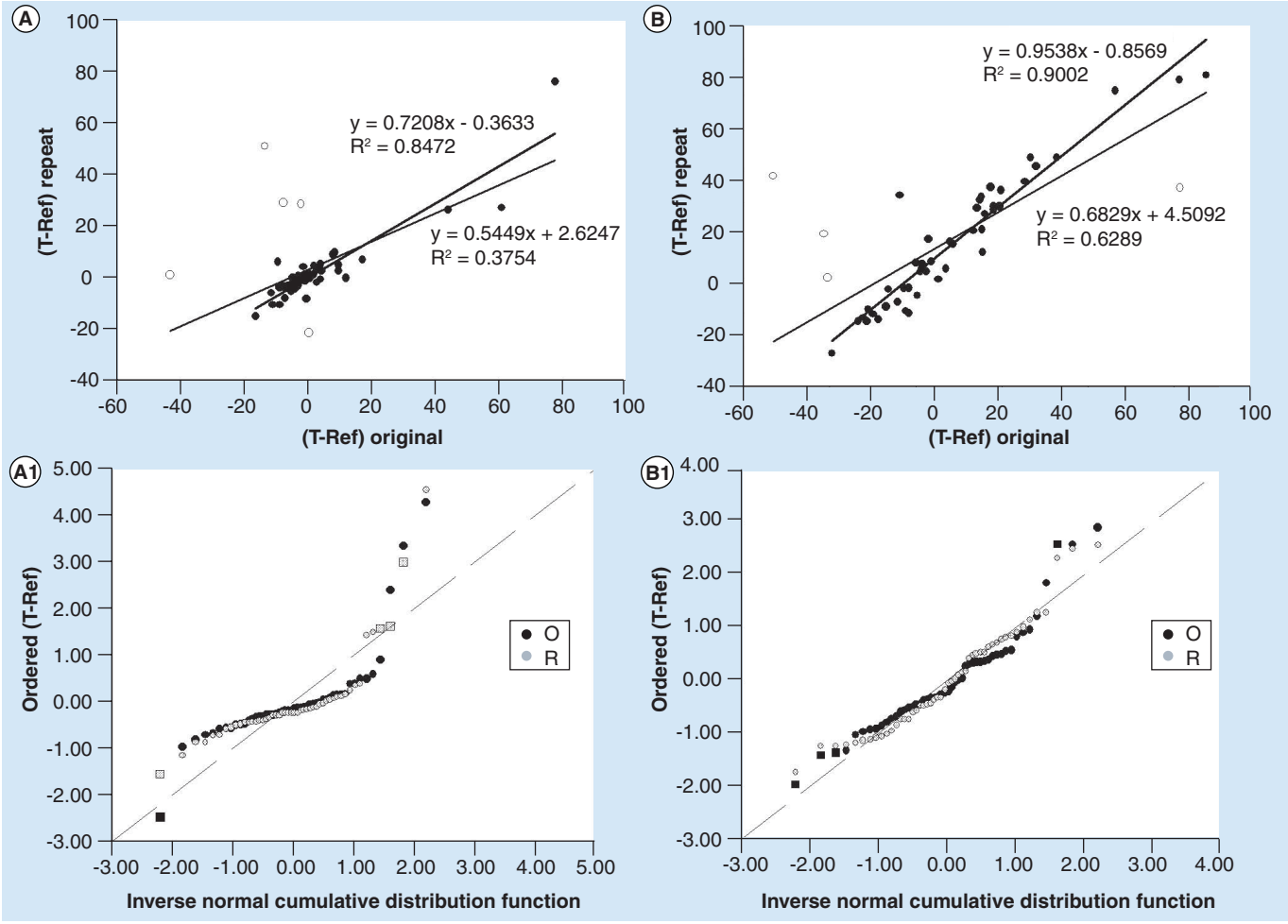


Figure 2. Regression and probability plots for spironolactone (A and A1) and canrenone (B and B1) made on (T - Ref.) differences found for the original and repeat data pairs. Open circles in **(A)** and **(B)** are outliers, while squares (black or grey) in **(A1)** and **(B1)** are the correspondence of the outliers in the probability plots.
O: Original; R: Repeat; Ref. Reference; T: Tested.

higher than 0.96). For spironolactone, although the patterns for the original and repeat events are very similar, it seems that the collections of values are not described by a single normal distribution, as the slopes of the linear regressions in Figure 2A1 are 0.8432 and 0.8533, respectively. This indicates that the calculated standard deviation is larger than the true spread of the experimental data.

Through elimination of the outliers (open circles in Figure 2A & B), the correlation coefficient (r_{xy}) increases from 0.61 to 0.92 and the slope from 0.545 to 0.721 for spironolactone (five outliers eliminated), while for canrenone r_{xy} jumps from 0.79 to 0.95 and the slope from 0.683 to 0.954 (four outliers eliminated). The statistical comparison of slopes to the unit value became positive for canrenone ($p > 0.3402$), but is still negative for spironolactone. Both intercepts are not significantly different from zero ($p > 0.6575$ for spironolactone and $p > 0.3781$ for canrenone). These outliers are corresponding to points situated at low and high values in the probability plots, representing rare events coming from the tail of the normal distributions (illustrated by squares in Figure 2A1 & B1). In addition, the use of ordinary linear regression imposes some other assumptions (normal distribution, the variable x is assumed to be error-free and constant error associated with y variable). These conditions are not always fulfilled and, as a consequence, contradictory results may be obtained.

Data resulting from the linear regression approach are suggestive and apparently easy to interpret through a simple visual inspection. However, the final results are seriously affected by the presence of outliers. In this respect, objective tools for assessment of the outlier status are undoubtedly needed to support their elimination prior to data processing.

Conclusion

Reproducibility data about the assay of spironolactone and canrenone in plasma samples resulting from a two-stage ISR study were discussed. Absolute concentration values were treated according to the Bland–Altman approach, leading to the conclusion that over the short-term delay the analytical method behaves in a reproducible manner. However, the reanalysis of study samples after a much longer period leads to the conclusion that data reproducibility is poor, at least concerning the analyte spironolactone. The explanation deals with the progressive transformation of spironolactone to canrenone in plasma, even on storage under frozen conditions at -40°C. The long-term period assessment was clearly placed outside the window of the long-term stability of 3 months established for the analytical method. Hence the results from these statistical evaluations should not have come as a surprise as from recommendations for bioequivalence studies the data must be generated within the period of established long-term stability for the method.

As the bioequivalence between two pharmaceutical formulations is decided based on differences between concentrations of the analytes in samples withdrawn from each volunteer at given sampling moments after administration of the tested and reference products, it seemed reasonable to asseses the method's reproducibility based on concentration differences between administration phases at given sampling moments obtained from volunteers during the original and repeat events. Differences between concentration values from the two clinical phases and different sampling moments failed to fulfil acceptance criteria after the application of the Bland–Altman procedure. Alternative statistical evaluation tools, namely the summary statistics, the paired t-test, the **box and whisker plot**, the Wilcoxon sign rank-sum test and linear regression were used for testing the similarity between the paired set of data. The linear regression and the statistical comparison of the regression parameters, although visually persuasive, are strongly affected by outliers, easily inducing misinterpretation of the results. It is still debatable if ISR should concentrate on the absolute reproducibility of the analytical results or should make emphasis more on random errors, as long as systematic biases do not substantially affect the final statement about the bioequivalence between two pharmaceutical formulations.

Future perspective

Although the need for evaluation of the reproducibility in bioanalysis is stringent, interpretation of data produced through ISR should be attentively considered in conjunction with the quality characteristics of the analytical method determined through validation and study completion. Alternative statistic models should be used for data assessment, especially in case of a failure towards the compliance to the general acceptance criteria (i.e., those derived from the Bland–Altman approach). A gradual shift from a statistical evaluation mode to another has to be developed to assist bioanalysts when experiencing reproducibility problems issuing from ISR data.

Key Term

Box and whisker plot: Graphical method of displaying the important characteristics of a set of observations. The display is based on the five-number summary of the data with the 'box' part covering the interquartile range, and the 'whiskers' extending to include all except outside observations, which are indicated separately.

Executive summary

- The reproducibility of a bioanalytical method used for the assay of spironolactone and its active metabolite canrenone in plasma samples intended for bioequivalence was tested by means of incurred sample reanalysis (ISR).
- ISR was made over two stages; one stage was performed close to the bioequivalence study completion moment, the second stage after a period substantially higher compared with the period assessed in the validation of the method for sample stability over long-term storage.
- For the short-term ISR study, the Bland–Altman procedure was applied on both log transformed data differences and mean normalized differences. Procedures produce similar conclusions, and sustain the reproducibility of the method through the fulfilment of the principal acceptance criteria.
- For the long-term ISR study, the same approaches led to the following conclusions: repeat values obtained for spironolactone are negatively biased compared with the original ones, ISR data failing to fulfil the acceptance criteria; for canrenone, despite an observable positive systematic bias, results are compliant to the principal acceptance criteria. Explanation of facts is based on the conversion of spironolactone to canrenone during long-term storage at -40°C. As plasma concentration levels of spironolactone are lower compared with canrenone ones, the inter-conversion was unable to push the concentration values of the metabolite outside the allowed limits.
- Alternatively, the reproducibility of the method may be assessed through comparison of paired data (original and repeat events): consisting in the differences between absolute concentration values determined at a given sampling moment in volunteer's samples resulting after administration of the tested (T) and reference (Ref.) formulations.
- Comparison between datasets representing the difference between concentrations in samples issued in the two administration stages was made through summary statistics, paired t-test, box and whisker plots, the Wilcoxon sign rank-sum test and linear regression.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Shah VP. The history of bioanalytical method validation and regulation: evolution of a guidance document on bioanalytical methods validation. *AAPS J.* 9(1), E43–E47 (2007).

2 US FDA. Guidance for Industry: Bioanalytical Method Validation. US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine (2001).

3 European Medicines Agency. *Guideline on Validation of Bioanalytical Methods*. EMA, Committee for Medicinal Products for Human Use (CHMP), London, UK (2009).

4 Viswanathan CT, Bansal S, Booth B *et al.* Workshop/conference report – Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *AAPS J.* 9(1), E30–E42 (2007).

5 Fast DM, Kelley M, Viswanathan CT *et al.* Workshop report and follow-up – AAPS workshop on current topics in GLP bioanalysis: assay reproducibility for incurred samples – implication of Crystal City recommendations. *AAPS J.* 11(2), 238–241 (2009).

6 Timmerman P, Luedtke S, van Amsterdam P *et al.* Incurred sample reproducibility: views and recommendations by the European Bioanalysis Forum. *Bioanalysis* 1(6), 1049–1056 (2009).

7 Savoie N, Garofolo F, van Amsterdam P *et al.* 2009 White Paper on recent issues in regulated bioanalysis from the 3rd calibration and validation group workshop. *Bioanalysis* 2(1), 53–68 (2010).

8 Blume H, Brendel E, Brudny-Klöppel M *et al.* Workshop/conference report on EMA draft guideline on validation of bioanalytical methods. *Eur. J. Pharm. Sci.* 42(3), 300–305 (2011).

9 Gilbert MT, Barinov-Colligon I, Miksic JR. Crossvalidation of bioanalytical methods between laboratories. *J. Pharm. Biomed. Anal.* 13(4–5), 385–394 (1995).

10 Thway MT, Ma M, Lee J *et al.* Experimental and statistic approaches in method cross-validation to support pharmacokinetic decisions. *J. Pharm. Biomed. Anal.* 49(3), 613–618 (2009).

11 James CA, Breda M, Frigerio E. Bioanalytical method validation, a risk based approach? *J. Pharm. Biomed. Anal.* 35(4), 887–893 (2004).

12 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327, 307–310 (1986).

13 Bland JM, Altman DG. Comparing methods of measurements: why plotting difference against standard method is misleading. *Lancet* 346, 1085–1087 (1995).

14 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* 8, 135–160 (1999).

15 Altman DJ, Bland JM. Commentary on quantifying agreement between two methods of measurement. *Clin. Chem.* 48(5), 801–802 (2002).

16 Rocci ML Jr, Viswanath D, Haughey DB *et al.* Confirmatory reanalysis of incurred bioanalytical samples. *AAPS J.* 9(3), E336–E343 (2007).

17 Lytle FE, Julian RK, Tabert AM. Incurred sample reanalysis: enhancing the Bland–Altman approach with tolerance intervals. *Bioanalysis* 1(4), 705–714 (2009).

▪ **Refers to the Bland–Altman approach based on mean normalized differences and the use of tolerance limits to evaluate reproducibility.**

18 Thway TM, Macaraeg CR, Calamba D *et al.* Bionalytical method requirements and statistical considerations in incurred sample reanalysis for macromolecules. *Bioanalysis* 2(9), 1587–1596 (2010).

19 Larson M, Han F. Determination of rifalazil in dog plasma by liquid–liquid extraction and LC–MS/MS: quality assessment by incurred sample analysis. *J. Pharm. Biomed. Anal.* 45(4), 616–624 (2007).

20 Brockman AH, Hatsis P, Paton P *et al.* Impact of differential recovery in bioanalysis: the example of bortezomib in whole blood. *Anal. Chem.* 79(4), 1599–1603 (2007).

21 Tan A, Jin W, Deng F *et al.* Bioanalytical method development and validation using incurred samples – simultaneous quantitation of ramipril and ramiprilat in human EDTA plasma by LC–MS/MS. *J. Chromatogr. B* 877(29), 3673–3680 (2009).

22 Silvestro L, Gheorghe MC, Tarcomnicu I *et al.* Development and validation of an HPLC–MS/MS method to determine clopidogrel in human plasma. Use of incurred samples to test back-conversion. *J. Chromatogr. B* 878 (30), 3134–3142 (2010).

23 Udrescu S, Sora ID, Albu F *et al.* Large volume injection of 1-octanol as sample diluent in reversed phase liquid chromatography: application in bioanalysis for assaying of indapamide in whole blood. *J. Pharm. Biomed. Anal.* 54(5), 1163–1172 (2011).

24 Jemal M, Ouyang Z, Powell ML. A strategy for a post-method-validation use of incurred biological samples for establishing the acceptability of a liquid chromatography/tandem mass-spectrometric method for quantitation of drugs in biological samples. *Rapid Commun. Mass Spectrom.* 16(16), 1538–1547 (2002).

25 Eastwood BJ, Farmen MW, Iversen PW *et al.* The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration–response assays and estimation by replicate – experiment studies. *J. Biomol. Screen.* 11(3), 253–261 (2006).

26 Petersen PH, Stöckl D, Blaabjerg O *et al.* Graphical interpretation of analytical data from comparison of a field method with reference method by use of difference plots. *Clin. Chem.* 43(11), 2039–2046 (1997).

27 Tokumura T, Muraoka A, Masutomi T, Machida Y. Stability of spironolactone in rat plasma: strict temperature control of blood and plasma samples is required in rat pharmacokinetic studies. *Biol. Pharm. Bull.* 28(6), 1126–1128 (2005).

28 Miller JN, Miller JC. In: *Statistics and Chemometrics for Analytical Chemistry*. 39–69; 107–135, 150–159, Pearson Education Ltd, Harlow, UK (2005).

▪▪ **Provides alternative statistical evaluation approaches for paired datasets.**

29 Rowe P. In: *Essentials Statistics for the Pharmaceutical Sciences*. 133–144; 169–192; 221–242, John Wiley & Sons Ltd, Chichester, UK (2007).

30 Chap Le T. In: *Introductory Biostatistics*. 246–263; 282–292, John Wiley & Sons Inc., NJ, USA (2003).