1 14 July 2016
2 EMA/CVMP/SWP/735325/2012
3 Committee for Medicinal Products for Veterinary Use (CVMP)

4 Guideline on approach towards harmonisation of 5 withdrawal periods

6 Draft

Draft agreed by Safety Working Party (SWP-V)	May 2016
Adopted by CVMP for release for consultation	14 July 2016
Start of public consultation	25 July 2016
End of consultation (deadline for comments)	31 January 2017

7

8 This guideline replaces the 'Note for guidance on approach towards harmonisation of withdrawal
9 periods' (EMEA/CVMP/036/95).

10

Comments should be provided using this template. The completed comments form should be sent to vet-guidelines@ema.europa.eu

11

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Introductory note on updates introduced in March 2016

In January 2014 the CVMP published a concept paper (EMA/CVMP/SWP/285070/2013) proposing a revision of the Note for guidance: Approach Towards Harmonisation of Withdrawal Periods, in order to look again at the approach used for considering residues present at levels below the limit of quantification (LOQ). The concept paper noted that the original Note for guidance recommends that a value of half of the limit of quantification should be applied to data points below the limit of quantification, but that since publication of the Note for guidance, more sophisticated methods for dealing with levels below the limit of quantification have become available, such as the maximum likelihood approach (i.e. determining the depletion curve that would maximise the likelihood of the observed data).

Following the receipt of comments on the concept paper, the SWP undertook work comparing the withdrawal periods calculated using different approaches for dealing with values below the LOQ. This work indicated that the current method (assigning values below the LOQ to half the LOQ) provides results that are comparable to those obtained using the maximum likelihood approach and also to using data 'as measured'. This supports the view that the current approach remains appropriate and that there is little to be gained by moving to an alternative. The CVMP therefore concluded that the existing approach for the treatment of values below the LOQ should remain in place. However, it should be noted that VICH GL49 recommends methods for determining the LOQ that are likely to make this issue less of a problem (as LOQs are likely to be < ½ MRL).

The work undertaken by the SWP in order to arrive at this conclusion is briefly described in the following sections of this introductory note.

In addition to adding this introductory note, the opportunity has been taken to add a number of clarifications to the guidance, to update references where appropriate (references to Regulation 2377/90 have been replaced with references to Regulation 470/2009, references to VICH GL48 & 49 have been added, reference to the guideline on injection site residues and the Draft reflection paper on injection site residues: considerations for risk assessment and residues surveillance have been added) and to bring the document in line with the EMA's current structure for guidelines. The clarifications added are:

Section 4.2: text added at beginning of section providing guidance on when it may not be appropriate to use the statistical approach.

Section 4.2: text added to end of section providing examples of how different factors might influence the size of the safety span

Section 6.5: text added highlighting that there should be a strong causal justification for removing values considered to be statistical outliers

Section 6.6: this section on the possibility of combining data sets has been added

Section 6.7: this section on the possibility of overriding a study has been added

Annex D: the final paragraphs, relating to specific problems concerning milk, have been deleted and replaced with a reference to the CVMP Note for guidance for the determination of withdrawal periods for milk.

Comparisons of different approaches for dealing with values below the LOQ

In a first step the SWP compared the following approaches:

(i) Omitting values below the LOQ;

(ii) Assigning a value of half the LOQ to values recorded as below the LOQ;

(iii) Using the maximum likelihood approach (i.e. the regression parameters were determined in such a way that the likelihood of observing the given values above the LOQ and the given frequency of values below the LOQ is maximised).

The results provided for liver in Annex A of the Note for guidance were used as the starting point from which to generate simulated data sets (derived based on the intercept, slope and standard deviation of the original data). Withdrawal periods were then derived from the (log transformed) simulated data sets either (i) omitting values below the LOQ, (ii) using values of half the LOQ when recorded values were below the LOQ, or (iii) using regression parameters based on the maximum likelihood approach. The original data set was considered to represent reality and to yield the 'true' withdrawal period, i.e. to yield a withdrawal period at the end of which 95% of all residue concentrations were, at most, as high as the MRL.

In principle, if a sufficient number of simulated data sets is sampled and withdrawal periods derived, then the frequency of withdrawal periods that are shorter than the 'true' withdrawal period should be 5% as, in line with the guideline, withdrawal periods should be derived in such a ways as to provide 95% confidence that they are not too short.

When withdrawal periods were derived treating values below the LOQ as described above, the following results were obtained:

(i) when values below the LOQ were omitted 1.3% of estimated withdrawal periods were at most as long as the 'true' withdrawal period (i.e. 98.7% were longer);

(ii) when values below the LOQ were replaced by a value of half the LOQ 5.6% of estimated withdrawal periods were at most as long as the 'true' withdrawal period (i.e. 94.4% were longer);

(iii) when the maximum likelihood approach was used to replace values below the LOQ 6.8% of estimated withdrawal periods were at most as long as the "true" withdrawal period (i.e. 93.2% were longer).

In this example, the method currently used in the EU came closest to the 5% value, with the maximum likelihood approach being almost as good.

The above exercise was then repeated using a further four real data sets and the withdrawal periods of the simulated data sets derived treating values below the LOQ, as described above. In addition, a fourth approach was used in which withdrawal periods were derived by using the values recorded for values below the LOQ ('as measured' values).

For each of the four approaches withdrawal periods for the simulated data sets were derived using three different assigned LOQs (LOQ assigned so that the expected percentage of values below the LOQ was 5%, 10% or 20%) and using MRLs set to either twice the LOQ or 5 times the LOQ, resulting in six different combinations of assigned LOQ and MRL for each data set.

90 The results are summarised in the table below.

			Approach for dealing with values below LOQ (BLOQ)			
Data set	%BLOQ	MRL	Omit	LOQ/2	As measured	Max Likelihood
A	5%	5 x LOQ	2.8	5.1	5.3	5.6
		10 x LOQ	1.9	6.4	5.4	5.5
	10%	5 x LOQ	2.2	4.5	4.7	4.8
		10 x LOQ	1.2	4.9	3.8	3.8
	20%	5 x LOQ	1.8	3.3	3.8	3.7
		10 x LOQ	1.0	4.7	3.6	3.7
B	5%	5 x LOQ	3.1	3.8	5.6	5.6
		10 x LOQ	2.1	5.9	5.1	5.4
	10%	5 x LOQ	2.4	2.8	4.3	4.2
		10 x LOQ	1.5	4.6	3.6	3.9
	20%	5 x LOQ	3.0	2.6	4.7	4.6
		10 x LOQ	1.6	4.6	3.8	4.0
C	5%	5 x LOQ	7.8	2.1	6.8	6.8
		10 x LOQ	2.6	3.0	5.6	5.8
	10%	5 x LOQ	7.6	1.2	5.7	5.6
		10 x LOQ	1.6	2.3	4.3	4.2
	20%	5 x LOQ	11.7	1.2	6.7	6.6
		10 x LOQ	1.6	1.6	3.9	3.9
D	5%	5 x LOQ	2.7	4.8	5.2	5.4
		10 x LOQ	1.6	5.8	4.5	4.8
	10%	5 x LOQ	2.1	3.9	4.2	4.2
		10 x LOQ	1.4	5.1	3.9	4.1
	20%	5 x LOQ	1.9	2.8	3.7	3.7
		10 x LOQ	1.2	4.3	3.3	3.3

91 The following observations can be made from the above table.
92

93 Omitting levels below the LOQ never came closest to yielding the desired frequency of 5% of
94 withdrawal periods shorter than the 'true' withdrawal period. In most cases it was the most
95 conservative method. This may be because omitting very low recorded residue levels will tend to make
96 the regression line less steep.

97 Using 'as measured' values for values below the LOQ yielded good results. However, it should be noted
98 that in the simulation constant variability of (log-transformed) data was assumed. With real data sets
99 higher variability is often seen at low residue levels (as described by the Horwitz equation). Therefore,
100 the apparent appropriateness of this method could be an artifact of the simulation's simplicity. Another
101 potential difficulty with this approach is that measurements below the limit of quantification are often
102 not reported.

103 Assigning values below the LOQ as half the LOQ and the maximum likelihood approach yielded
104 similarly appropriate results in most cases – withdrawal periods were generally similarly distributed,
105 and the fraction of withdrawal periods at most as long as the 'true' withdrawal period were similar.
106 However, for one data set (data set C) the maximum likelihood approach does appear to have yielded
107 better results.

108 Overall, the 'as observed' approach, the half LOQ approach and the maximum likelihood approach can
109 be considered to have yielded similar results, with the percentage of withdrawal periods that are too
110 short ranging from approximately 3% to less than 7%, corresponding to a confidence more than 93%
111 to approximately 97%.

It is acknowledged that the above investigation is limited and that further work could be undertaken to further explore different approaches for dealing with values below the LOQ and for investigating whether all assumptions used in derivation of withdrawal periods are supported. In reality it is likely that there is not one single method that will be optimal for dealing with all data sets. Ideally, software would be developed that would automatically identify and apply the most appropriate approach. However, the development of such software would be a very substantial undertaking. VICH GL 49 (adopted by CVMP, March 2011) recommends that the LOQ for an analytical method should be estimated as the mean of 20 control samples plus 6-10 times the standard deviation (SD), and then confirmed, or be based on the ability of the method and the instrumentation used to detect and quantify a specific analyte in a specific matrix (see Annexes 1 & 2 of GL49). Before GL49 was adopted, the LOQ was routinely determined as 0.5 x MRL, leading to many results being reported as 'below LOQ' (<LOQ or BLOQ). With the guideline-recommended method of determining the LOQ, it is foreseen that there will be fewer data <LOQ, as the difference between LOQ and MRL would usually be greater than that between 0.5 x MRL and MRL. This should lead to fewer issues around which values to use, as the depletion curve would be better described.

Guideline on approach towards harmonisation of withdrawal periods

Table of contents

Introductory note on updates introduced in March 2016 2

Executive summary 7

1. Introduction (background) 7

2. Scope 8

3. Legal basis 8

STATISTICAL APPROACH TO THE ESTABLISHMENT OF WITHDRAWAL PERIODS 8

4. General considerations 8

4.1. Statistical approach 8

4.1.1. Calculation model 8

4.1.2. Data base 9

4.1.3. Linear regression analysis assumptions 9

4.1.4. Estimation of withdrawal periods by regression analysis 11

4.2. Possible alternative approach 12

4.3. Injection site residues 13

5. Example for the statistical analysis of residue data 14

6. Discussion on the regression analysis 22

6.1. To what extent a departure from the regression assumptions may be acceptable? 22

6.2. Withdrawal periods should be set by interpolation and not by extrapolation. 22

6.3. Should the 95% or the 99% tolerance limit be applied? 22

6.4. Dealing with 'less than' values 23

6.5. Dealing with obvious outliers. 23

6.6. Combining data sets 23

6.7. The possibility of overriding one study with another 24

7. References 24

Annex A 26

Annex B 1 28

Annex B 2 32

Annex C 34

Annex D 36

Executive summary

The document originally published in 1997 as the CVMP Note for guidance: approach towards harmonisation of withdrawal periods, provides detailed guidance on how to establish withdrawal periods and was developed by the CVMP in order to provide a standardised approach for derivation of withdrawal periods within the European Union. Much of the document is focused on the statistical approach used by CVMP, but an alternative, for use in those cases where the data do not allow use of the statistical approach, is also described. The issue of withdrawal periods for substances with a 'No MRL required' classification is also addressed.

1. Introduction (background)

1. Even where Community MRLs have been established, similar products in various Member States may differ greatly with respect to the withdrawal periods established by national authorities.

2. The 1997 note for guidance enabled applicants and assessors from all member states to use the same approach for determining withdrawal periods (WPs), leading to fewer discrepancies between authorised WPs for the same product in different member states (MS). The same approach is also used in centralised and decentralised procedures.

3. The Committee considers that the statistical approach offers the greatest opportunity for harmonisation but recognises there are occasions when a simpler, more pragmatic approach is necessary and recommends the following:

New chemical entities

4. As residue depletion studies for the establishment of withdrawal periods should be conducted in accordance with Volume VIII of the Rules governing Medicinal Products in the European Community, and VICH GLs 48 and 49, data should be sufficiently adequate to use a statistical method.

5. Applicants should use the statistical software provided by the CVMP (found on the EMA website) in order to determine a suitable WP for their product(s). The underlying statistics for this software are described in the Annex to this Guideline.

Old chemical entities

6. In many cases, depletion studies could have been conducted before the publication of the requirements indicated in Volume VIII, or VICH GLs 48 and 49, so the data are insufficient to evaluate the withdrawal period using the recommended statistical method.

7. For this reason, an alternative method, which has been used successfully throughout the union for many years, has also been included; however, it should only be used where the statistical method(s) cannot be used.

The objective of the present paper is to provide guidance on how to establish withdrawal periods for edible tissues of food producing animals. This guideline does not address withdrawal periods in milk, for which guidance is provided in the CVMP Note for guidance for the determination of withdrawal periods for milk (EMEA/CVMP/473/98-FINAL).

Emphasis has been put on a statistical approach. As the method of first choice, linear regression technique is recommended. Data from an actual residue study were used to demonstrate the applicability of this recognized statistical technique. A step by step procedure is described which has

been drawn up with the FDA guideline (1, 2) as a basis. It is recommended in this paper to determine withdrawal periods at the time when the upper one-sided 95 % tolerance limit for the residue is below the MRL with 95% confidence. However, for comparison of approaches (cf. FDA), 99% tolerance limits with 95% confidence are also calculated.

2. Scope

This guideline describes a standardised approach for the determination of withdrawal periods within the European Union, focusing particularly on use of a statistical method but providing additional guidance on an alternative approach, for use in those cases where the data do not allow use of the statistical approach (i.e. where the statistical assumptions are not met).

In addition, the paper discusses the possible need for withdrawal periods for products containing substances for which a 'No MRL required' status has been established, as well as generic products.

3. Legal basis

In line with article 12.3 of Directive 2001/82/EC, marketing authorisation applications for veterinary medicinal products for use in food producing species must include an indication of the withdrawal period. Article 1.9 of the directive defines the withdrawal period as:

> *The period necessary between the last administration of the veterinary medicinal product to animals, under normal conditions of use and in accordance with the provisions of this Directive, and the production of foodstuffs from such animals, in order to protect public health by ensuring that such foodstuffs do not contain residues in quantities in excess of the maximum residue limits for active substances laid down pursuant to Regulation (EEC) No 2377/90.*

STATISTICAL APPROACH TO THE ESTABLISHMENT OF WITHDRAWAL PERIODS

4. General considerations

4.1. Statistical approach

4.1.1. Calculation model

The calculation model for the statistical determination of withdrawal periods is based on accepted pharmacokinetic principles. According to the pharmacokinetic compartment model, the relationship between drug concentration and time through all phases of absorption, distribution and elimination is usually described by multiexponential mathematical terms. However, the terminal elimination of a drug from tissues, the residue depletion, in most cases follows a one compartment model and is sufficiently described by one exponential term. The first order kinetic equation for this terminal elimination is:

$$C_t = C_o' \, e^{-kt}$$

C_t is the concentration at time t, C_o' is a pre-exponential term (fictitious concentration at t=0) and k is the elimination rate constant.

Linearity of the plot $\log_e C$ versus time indicates that the model for residue depletion is applicable and linear regression analysis of the logarithmic transformed data can be considered for the calculation of withdrawal periods.

4.1.2. Data base

Regression analysis requires data which are independent from each other. Normally, residue depletion data meet this assumption because they originate from individual animals. In cases of duplicate or triplicate measurements of samples the mean value of each sample has to be used for the calculation. To avoid biasing slope and intercept, each data point of the regression line should originate from the same number of repeated sample measurements. However, the effect of the analytical error on the final results, in most cases, is very small compared with the effect of animal to animal variability.

The FDA (1) recommends excluding from the calculation data observed as below the limit of detection. In the Committee's opinion, this approach biases the regression line. As the low concentrations are due to real empirical observations they should not be ignored.

Therefore, setting the data which are below the limit of detection or quantitation ('less than' values) to one-half of the respective limit is recommended. Alternatively, special procedures may be applied in order to estimate the expected values for missing data. Possible approaches are described by Helsel or Newman (11, 12).

When all or most of the reported data of a slaughter day are 'less than' values it should be considered to exclude the whole time point. However, it should be borne in mind that 3 time points are necessary to allow a meaningful regression analysis.

The numbers of animals to be used for residue depletion studies is specified in guideline VICH GL48: Studies to evaluate the metabolism and residue kinetics of veterinary drugs in food-producing animals: marker residue depletion studies to establish product withdrawal periods EMA/CVMP/VICH/463199/2009 (14 March 2011). There, depending on the animal species and type of depletion study, 4-10 animals per time point are recommended.

Remark: Usually, analytical values are reported as they are measured (uncorrected for recovery) with supporting data involving recovery experiments. Therefore, in these cases, a correction for recovery has to be carried out prior to any calculation of withdrawal periods.

4.1.3. Linear regression analysis assumptions

It is necessary for linear regression analysis that the following regression assumptions are valid:

- assumption of homogeneity of variances of the $\log_e$ -transformed data on each slaughter day,
- assumption of linearity of the $\log_e$-transformed data versus time,
- assumption of a normal distribution of the errors.

4.1.3.1 Homogeneity of variances

It should be confirmed that the variances of the $\log_e$-transformed concentrations of the different slaughter days are homogeneous.

Several tests are available. The FDA (1, 2) recommends Bartlett's test. Bartlett's test is said to be the most powerful test, but it is extremely sensitive to deviations from normality. Furthermore, the test should only be used, when each group numbers 5 or more. Equal sample sizes are not required (3).

Other commonly used tests for homogeneity of variances are Hartley's test and Cochran's test. Hartley's test can only be used if all groups are of the same size (3).

In the Committee's view, Cochran's test is the best choice. It is easier to perform than the test of Bartlett, and it uses more information than Hartley's test. Furthermore, it is not as sensitive to departures from normality as the test of Bartlett. Cochran's test may be used for data whose group sizes do not differ substantially by calculating the harmonic mean of the group sizes.

4.1.3.2 Log-linearity

Visual inspection of a plot of the data is often sufficient to assure that there is a useful linear relationship. Obvious deviations from linearity at early time points may indicate that the drug distribution processes have not yet ended. These time points should therefore be excluded. Deviations from linearity at late time points may be due to concentrations below the limit of detection. Depletion kinetics cannot be observed at these time points, and it is justified to exclude these data. It should, however, be borne in mind that all other time points have to be kept, unless there is a clear justification for their omission.

For statistical assurance of the linearity of the regression line an analysis of variances has to be performed (lack-of-fit test). The usual procedure is to compare the variation between group means and the regression line with the variation between animals within groups (see Section 5, Step 5).

An appropriate supplementation to the lack of fit test is the test of the significance of the quadratic time effect according to Mandel (10). The question is, whether a quadratic fit is better than the linear fit. The calculation procedure is described in Annex C of this paper.

4.1.3.3 Normality of errors

A good visual test is to plot the ordered residuals versus their cumulative frequency distribution on a normal probability scale. Residuals are the differences between the observed values and their expectations (i.e. the difference between the observed $\log_e$-transformed concentration and the value predicted by the regression line).

A straight line indicates that the observed distribution of residuals is consistent with the assumption of a normal distribution. In order to verify the results of the residual plot, the Shapiro-Wilk test can be applied. This test has been shown to be effective even if sample sizes are small (4).

The plot of the cumulative frequency distribution of the residuals can be used as a very sensitive test. Deviations from a straight line, indicating non-normality of the residuals, may be due to:

- deviations from normality of the $\log_e$-transformed residue concentrations within one or more slaughter groups,
- deviations from $\log_e$-linearity of the regression line,
- non-homogeneity of variances,
- outliers.

In the selected presentation of the data using standardized residuals (standardized by dividing by the residual error $s_{y.x}$), an outlier would have a value < –4 or > +4, indicating that the residual is 4 standard deviations off the regression line (see Fig. 1, 2).

4.1.4. Estimation of withdrawal periods by regression analysis

The withdrawal period should be estimated using the results of linear regression calculations. Withdrawal periods are determined at the time when the upper one-sided tolerance limit with a given confidence is below the MRL. If this time point does not make up a full day, the withdrawal period is to be rounded up to the next day.

The FDA (1, 2) recommends calculating the 99th percentile of the population with a 95% confidence level by a procedure which requires the non-central t-distribution.

The calculation of the one-sided upper tolerance limit (95% or 99%) with a 95% confidence according to K. Stange (5) is proposed in this paper. This method of calculation has comparable results (see Annex B) and is easier to perform since only the percentage points of the standardized normal distribution are required.

With the Stange equation one estimates (with a confidence of $1-\alpha$) the proportion of $1-\gamma$ of the population which at least is to be expected to be below the one-sided upper tolerance limit. The respective percentage points of the standardized normal distribution are $u_{1-\alpha}$ and $u_{1-\gamma}$ (e.g. for $1-\alpha$ = 0.95 is $u_{1-\alpha}$ = 1.6449, for $1-\gamma$ = 0.95 is $u_{1-\gamma}$ = 1.6449, and for $1-\gamma$ = 0.99 is $u_{1-\gamma}$ = 2.32635).

The equation published by K. Stange (5) is:

$$y = a + bx + k_T \, s_{y.x}$$

with

$$k_T = \frac{\sqrt{(2n - 4)}}{(2n - 4)^* - u_{1-\alpha}^2} \left[\sqrt{(2n - 4)^*} \; u_{1-\gamma} + u_{1-\alpha} \; W_n \right]$$

$$W_n = \sqrt{u_{1-\gamma}^2 + \left[(2n-4)^* - u_{1-\alpha}^2 \right] \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}} \right]}$$

$$S_{xx} = \sum x_i^2 - \frac{1}{n} (\sum x_i)^2$$

$s_{y.x}$ = residual error $()^* = (2n-5)$, according to Graf et al. (6)

A revised version of the Stange equation (using the term (2n–5) instead of (2n–4) in the three parentheses marked above by an asterisk) was published by Graf et al. in 1987 (6). The use of this equation results in slightly higher tolerance limits. According to Stange (5) the equation is valid for $n \approx 10$, whereas Graf et al. (6) restrict validity to $n \approx 20$.

A listing of data comparing the results of both equations to the results of the FDA procedure can be found in Annex B1 of this paper.

Remark: For reasons discussed below (see Section 6.3) the selection of the 95% tolerance limit with 95% confidence is preferred.

4.2. Possible alternative approach

The statistical approach should be used whenever a data set fulfils the minimum requirements for a statistical analysis. The statistical significance levels given in this guideline should be considered as recommendations, not as strict rules, in that any violation would not automatically trigger use of an alternative approach. A decision to not use a statistical approach should always be scientifically justified and based on statistical expert judgement.

The following statistical tests are reffered to: F-test, Chochrane test, Bartlett test, Shapiro-Wilk test, the most critical of which is the lack of fit test (F-test). Significant deviations from a straight line cannot be accepted for the model recommended in the guideline.

In many cases, the question of whether the statistical method can be used or not is dependent on the number of time points with a sufficient number of observations above the LOQ; the validation of the LOQ is therefore pivotal in this regard. The statistical method could probably be used in more situations where a lower LOQ is demonstrated.

Whenever data available do not permit the use of the statistical model, an alternative approach has to be considered in order to determine appropriate withdrawal periods.

A general recommendation for such a procedure cannot be provided. A specific approach depends on many parameters such as sample size, number, frequency and choice of slaughter timepoints, variability of the data, and analytical factors (e.g. level of the detection limit (LOD), stability of analytes during matrix processing).

One concept is the establishment of the withdrawal period at the time point where the concentrations of residues in all tissues for all animals are at or below the respective MRLs (13). However, when one

has determined that time point, the estimation of a safety span should be considered in order to compensate for the uncertainties mentioned above.

The value of a safety span depends on various, not easy to specify, factors which are decided by the study design, the quality of the data and finally by the pharmacokinetic properties of the active substance(s). As a result, an overall recommendation cannot be provided. An approximate guide for a safety span is likely to be a value of 10% - 30% of the time point when all observations are at or below the MRL. Alternatively, a safety span might be calculated from the tissue depletion half-life, possibly a value of 1-3 times $t_{1/2}$.

Examples of how certain factors might influence the size of the safety span:

- If, at the first time point at which residues are below the MRL, all values are below the LOQ, then a safety span of 10% may be acceptable.
- If there are long gaps between time points and if residue levels are already close to the MRL at the timepoint before the one at which they actually fall below the MRL, then a safety span of 10% may be appropriate.
- If there is high variability between animals at each timepoint then a safety span of 30% may be appropriate.
- The proximity of the residue value to the MRL should be taken into account and a safety span at the higher end of the standard range (i.e. a safety span of 30%) considered in those cases where the residue finding is at the MRL.
- If the first time point at which all residues are below the MRL is < 10 days, then a longer safety span should be used (17).

4.3. Injection site residues

When considering the establishment of withdrawal periods for parenterally administered drugs, it is important to take into account the residues of the intramuscular (IM) or subcutaneous (SC) injection site. The guideline on injection site residues (EMEA/CVMP/542/03-FINAL) specifically addresses this point. The reader is also referred to the CVMP Draft reflection paper on injection site residues: considerations for risk assessment and residue surveillance (EMA/CVMP/520190/2007-Rev.1).

5. Example for the statistical analysis of residue data

Data constructed from an empirical residue depletion study on cattle treated subcutaneously with a veterinary drug were used to demonstrate the applicability of the statistical model for the estimation of withdrawal periods. The residue data for the marker residue in the target tissues liver and fat are listed in Table 1 (see Annex A). An ADI of 35 µg per day for a 60 kg person has been assumed for the total residue. The MRLs for the marker residue have then been set at 30 µg/kg and 20 µg/kg for liver and fat, respectively.

Calculation procedure

Step 1: Inspection of the data (listed in Table 1, Annex A)

As discussed earlier, data below the limit of detection (i.e. 2 µg/kg) were set to one-half of the detection limit (i.e. 1.0 µg/kg).

For fat, the day 35 was excluded from calculation because of too many values below the detection limit (10 of 12 observations). Data for liver on day 35 were not available.

Step 2: Calculation of the linear regression parameters of the $\log_e$-transformed data

Table 2: Linear regression parameters

Parameter	**Liver**	**Fat**
Number of values *	n = 48	n = 48
Intercept	a = 5.64 ± 0.35	a = 5.84 ± 0.36
Slope	b = – 0.16 ± 0.02	b = – 0.17 ± 0.02
Correlation coefficient	r = – 0.7927	r = – 0.8026
Residual error	$s_{y.x}$ = 0.9930	$s_{y.x}$ = 1.0258

* excluded data: day 35 for fat (day 35 for liver: not assayed)

Step 3: Visual inspection of the regression line

Both the regression line for liver and the regression line for fat passed through all slaughter groups. No time points have to be excluded at the end or at the beginning of the line (see Fig. 3 and 4).

Step 4: Homogeneity of variances

Due to the amount of data given per group and due to the equal group sizes, it was possible to use all three tests discussed above. The equations and percentage points have been published in L. Sachs (3). The results of the tests are summarized in the Tables 3-5.

Table 3: Bartlett's test

Tissue	Test value	Degrees of freedom	Probability	Significance
liver	$\hat{\chi}^2$ = 4.24	df = 3	P > 0.05	n.s.
fat	$\hat{\chi}^2$ = 5.95	df = 3	P > 0.05	n.s.

n.s.: differences are not significant

Table 4: Cochran's test

Tissue	Test value	Degrees of freedom	Probability	Significance
liver	$\hat{G}$ max= 0.343	df_1 = 11 df_2 = 4	P > 0.05	n.s.
fat	$\hat{G}$ max= 0.442	df_1 = 11 df_2 = 4	P > 0.05	n.s.

n.s.: differences are not significant

Table 5: Hartley's test

Tissue	Test value	Degrees of freedom	Probability	Significance
liver	$\hat{F}$ max=3.46	df_1 = 4 df_2 = 11	P>0.05	n.s.
fat	$\hat{F}$ max=4.68	df_1 = 4 df_2 = 11	P>0.05	n.s.

n.s.: differences are not significant

Conclusion: The variances of the $\log_e$-transformed data at each time point are homogeneous.

Step 5: Analysis of variances (showing lack of fit) according to L. Sachs (3)

The ratio

$$\hat{F} = \frac{\text{MS between group means and the regression line}}{\text{MS within groups}}$$

was calculated and compared to the 5% percentage point of the F-distribution. Generally, a significant ratio indicates that the $\log_e$-linear model appears to be inadequate.

Table 6: ANOVA table for liver

Source of variation	**Degrees of freedom**	**Sum of square (SS)**	**Mean square (MS=SS/df)**
Between group means and the regression line	2	0.784	0.3919
Within groups (departure of y-values from their group mean)	44	44.573	1.0130
$\hat{F}$ (test) = 0.3869 (df_1= 2, df_2= 44) P>0.05 n.s.			

n.s.: no significant deviation from linearity

Table 7: ANOVA table for fat

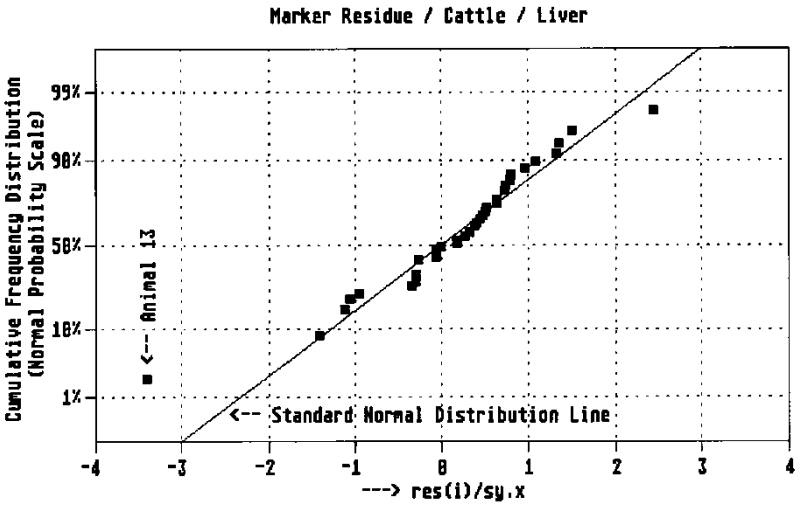
Source of variation	**Degrees of freedom**	**Sum of square (SS)**	**Mean square (MS= SS/df)**
Between group means and the regression line	2	6.240	3.1199
Within groups (departure of y-values from their group mean)	44	42.165	0.9583
$\hat{F}$ (test) = 3.2557 (df_1= 2, df_2= 44) 0.05> P>0.025 n.s. *			

* Potential deviation from linearity emerges.

Conclusion: In any case, the assumption of linearity of the $\log_e$-transformed data versus time can be upheld for liver. In the case of fat, a potential deviation from linearity emerges. A critical re-inspection of the plotted data (Fig. 4) suggests that day 7 may possibly belong to an earlier phase of residue depletion. Excluding day 7 from calculation might therefore be taken into account. This approach was not followed up here because the linearity assumption was not seriously violated.

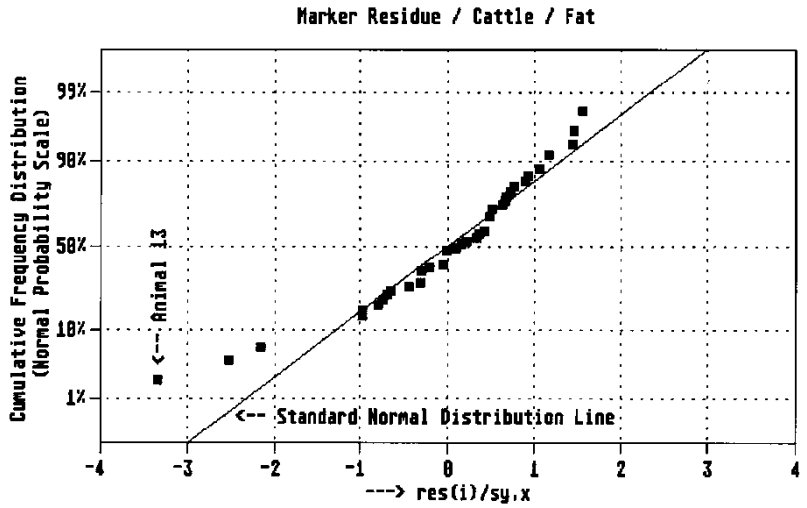
Step 6: Calculation of residuals and plot of cumulative frequency distribution according to the recommendation of the FDA 1983 (2)

The plots for the ordered residuals (standardized by the residual error $s_{y.x}$) versus their cumulative frequency on a normal probability scale are shown in Figure 1 (liver) and Figure 2 (fat).



444

445 Fig. 1: Cumulative frequency distribution of residuals for liver



446

447 Fig. 2: Cumulative frequency distribution of residuals for fat

448 Conclusion: Fat shows a marked departure from the straight line at the negative end of this line. The
449 value which deviates most belongs to the animal numbered 13. The plot for liver as well, shows that
450 the sample of animal 13 deviates from the standard normal distribution line. This is a possible
451 indication that the residue data of animal 13 tend to be outliers.

452 In order to verify the results of the residual plot, the Shapiro-Wilk test for normality was performed
453 according to G. B. Wetherill (4). The coefficients required for calculation of the test value $\hat{W}$ were
454 taken from Table C7 (see (4), pp. 378 - 379) and compared to the percentage points for the Shapiro-
455 Wilk-test, published in Table C8 (see (4) p. 380). The assumption of a normal distribution (in this case
456 a normal distribution of the errors) holds as long as the test value $\hat{W}$ exceeds the 10% percentage
457 point for the given sample size.

458

Table 8: Shapiro-Wilk test

Tissue	Test value	n	Probability	Significance
Liver	$\hat{W}$ = 0.960	48	P > 0.10	n.s.
Fat	$\hat{W}$ = 0.922	48	P < 0.01	*
Fat (animal 13 excl.)	$\hat{W}$ = 0.955	47	P > 0.10	n.s.

n.s.: No significant deviation from normality; * Significant deviation from normality

Conclusion: No deviation from normality could be observed for liver. For fat, there was a significant deviation of the errors from normality when testing all fat samples. As discussed above, the sample 13 may possibly be seen as outlier. Excluding animal 13 from calculation for fat, the distribution returned to normality.

Step 7: Calculation of the one-sided 95% and 99% upper tolerance limits (both with a 95% confidence level) according to K. Stange (5):

The numerical values are summarized in Table 9 and 10. Plots of withdrawal period calculations for liver and fat are shown in Figures 3 and 4.

Table 9: Results for liver (full data set, including animal 13):

Days post dose	Statistical tolerance limits with 95% confidence	
	95% Tolerance limit (µg/kg)	99% Tolerance limit (µg/kg)
26	35.7	77.9
27	30.9	67.4
28	26.8*	58.3
29	23.3	50.5
30	20.3	43.7
31	17.6	38.0
32	15.3	33.0
33	13.4	28.7*

* below the MRL (30 µg/kg) for liver

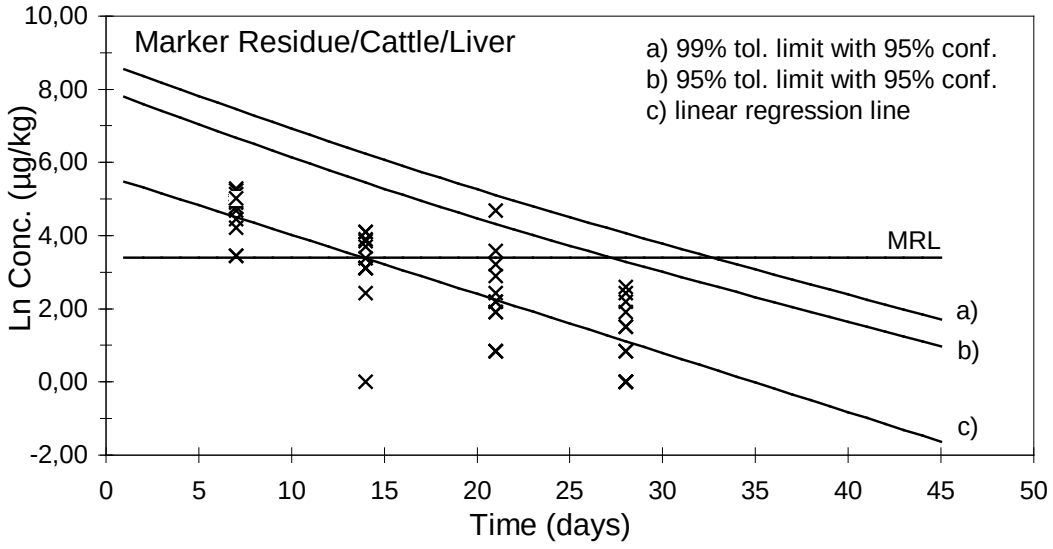
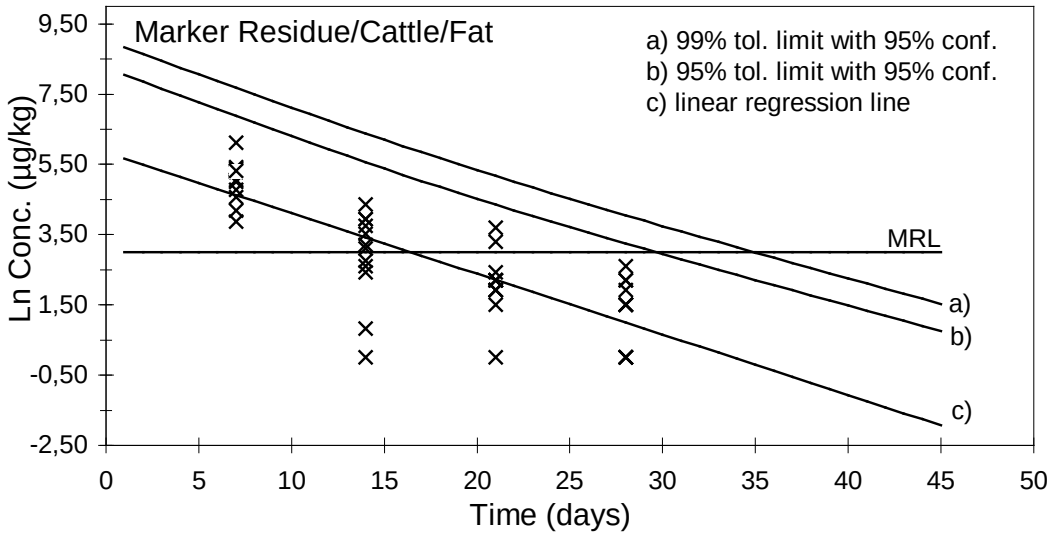


Fig. 3: Plot of withdrawal period calculation for liver

473 **Table 10:** Results for fat (full data set, including animal 13):

Days post dose	**Statistical tolerance limits with 95% confidence**	
	95% Tolerance limit (µg/kg)	99% Tolerance limit (µg/kg)
26	35.1	78.6
27	30.1	67.2
28	25.8	57.5
29	22.2	49.3
30	19.1*	42.3
31	16.4	36.3
32	14.2	31.2
33	12.2	26.8
34	10.5	23.1
35	9.1	19.9*
36		17.2

474 * below the MRL (20 µg/kg) for fat



475

476 Fig. 4: Plot of withdrawal period calculation for fat

477 The MRLs for the target tissues liver and fat are 30 µg/kg and 20 µg/kg, respectively. The time points
478 when the residues in fat and liver dropped below their MRLs are summarized in Table 11.

479 **Table 11:** Withdrawal periods obtained for the full data set including animal 13

Withdrawal times obtained from	**Liver**	**Fat**
95% tolerance limit (95% conf.)	28 days	30 days
99% tolerance limit (95% conf.)	33 days	35 days

480

Re-evaluation of data excluding animal 13

Table 12: Test results (excluding 13)

	Liver	**Fat**
Bartlett's test	0.05 > P > 0.025	P > 0.05
Cochran's test	P > 0.05	P > 0.05
Lack of fit test	P > 0.05	P > 0.05
Shapiro-Wilk test	P > 0.10	P > 0.10

The regression assumptions are not seriously violated.

Taking into account MRLs of 30 µg/kg and 20 µg/kg for liver and fat, respectively, the withdrawal times listed below were estimated:

Table 13: Withdrawal periods obtained (excluding 13)

Withdrawal times obtained from	**Liver**	**Fat**
95% tolerance limit (95% conf.)	26 days	29 days
99% tolerance limit (95% conf.)	31 days	33 days

Step 8: Estimation of the withdrawal period for the injection site (using an alternative approach)

In the example discussed here, the withdrawal periods estimated in Step 7 were based on the MRLs for the target tissues fat and liver. An MRL for muscle was not established for the drug under review. Therefore, the withdrawal period for injection site residues has to be calculated on the basis of the ADI being 35 µg (per day for a 60 kg person) for the total residue (listed in Table 1, Annex A).

It has to be shown that the ADI is not exceeded when the usual food package (0.5 kg) includes 0.3 kg injection site (instead of 'normal' muscle). In some cases, the CVMP will have set an ISRRV, which can be used as a surrogate for the muscle MRL for injection sites only (18).

For this purpose, marker residue concentrations from Table 1 were converted to total residues according to the average ratios marker/total (0.3 for liver, fat and kidney, and 0.6 for injection site muscle), determined in a total residue depletion study. The daily intake of the total residue from each tissue type was calculated using the standard food consumption figures (300 g injection site, 100 g liver, 50 g kidney and 50 g fat). In other words, the total residue in the 0.5 kg food package was determined for each slaughter day by using the following equation:

$$RI = (c_L \times F_L / R_L) + (c_K \times F_K / R_K) + (c_F \times F_F / R_F) + (c_M \times F_M / R_M)$$

RI = residue intake (µg)
c = concentration of the marker residue (µg/kg)
F = food consumption figures (0.3 kg muscle, 0.1 kg liver, 0.05 kg kidney, 0.05 kg fat)
R = ratio marker residue vs. total residue
(to be applied when the ADI refers to the total residues)
Indices L, K, F, M = liver, kidney, fat and muscle (here injection site)

Day 28 was not excluded from calculation even though there were only 2 values (out of 12) above the limit of detection for the injection site. However, day 35 was excluded because data for liver and

kidney were not available. Data below the limit of detection were set to one-half of the limit of detection. The results of this calculation are listed in the last column of Table 1 (Annex A).

As residue depletion from the injection site was rather erratic (high animal to animal variation) the statistical requirements for regression analysis were not met by these data for the daily dietary residue intake. The data revealed a significant deviation from normality and the homogeneity of variances was slightly violated.

Table 14: Test results

Edible portion	
Bartlett's test	0.05> P> 0.025 *
Lack of fit test	P> 0.05 n.s.
Shapiro-Wilk test	0.05> P> 0.02 **

n.s.: no significant deviation from linearity

* potential non-homogeneity of variances

** significant deviation from normality

Furthermore, the tolerance limits crossed the ADI-line far after the time range when data for the total residue intake were available (95% tolerance limit: day 35, 99% tolerance limit: day 42). Since the time period between day 28 and day 35/42 was not covered by data and since the regression assumptions were not met, the statistical approach of setting a withdrawal period seemed to be inadequate.

Therefore, an alternative approach was applied:

Inspection of the data for the daily dietary residue intake (Table 1) showed that on day 28 the highest individual residue amount (calculated as 32.3 µg) was just below the ADI being 35 µg. In order to account for the high variability of the residue data, especially the variability of the injection site data, a safety span has to be added to the depletion time of 28 days. A safety span of 7 days can be seen as appropriate. This safety span corresponds to 25% of the 28 day depletion time. The alternative approach would then result in a withdrawal period of 35 days.

On the whole, it should be noted here that any alternative approach is of course rather subjective and depends on the significance given to specific aspects of the information available.

Remark: The final withdrawal period has to be set in a way that the residues in all target tissues drop below their specific MRLs and ISRRVs, and, in addition, that the amount of residues in the edible portion drops below the ADI. This means, that the longest withdrawal period has to be selected in order to be in full compliance with the MRLs, ISRRV and the ADI. In the example discussed here, the withdrawal times obtained from the statistical 95% tolerance limits for fat and liver residues were 30 and 28 days, respectively. However, the withdrawal period of 35 days derived for the injection site would determine the conclusive withdrawal period.

6. Discussion on the regression analysis

Data on residues in cattle liver and fat (constructed from real empirical data) were analysed by using a set of basic statistical tests in order to prove that linear regression analysis is an appropriate model for estimation of withdrawal periods. It was shown that assumptions on which the regression analysis is based could in principle be upheld when tested on these data. Only in the case of fat was the normality assumption violated (Shapiro-Wilk test). However, excluding one sample (which was suspected to be an outlier) the distribution of the fat data returned to a normal distribution.

The statistical procedure applied to these data revealed a number of problems associated with estimating withdrawal periods:

6.1. To what extent a departure from the regression assumptions may be acceptable?

The first general question is where to set the significance levels of the tests and to what extent a departure from the regression assumptions may be acceptable. Second, should these assumptions absolutely dictate whether the calculation model can be used or not?

In other words, one could be faced with a situation in which the data do not sufficiently satisfy the statistical assumptions. In this situation one has to decide whether the calculation procedure should be stopped, strictly according to the rules of statistics, or whether the calculation procedure may be continued under more investigative considerations. As long as the regression assumptions are not seriously violated, the tolerance limits might be used as a reference for an appropriate safety span. In our view, this pragmatic approach will at least provide rough orientation for a potential withdrawal period.

6.2. Withdrawal periods should be set by interpolation and not by extrapolation.

In some cases, the concentrations of the MRLs are close to the LOQ of the analytical method which has been used to measure these residues. As a consequence, data nearest the time point when the upper tolerance limit crosses the MRL-line are not available. It seems, therefore, inevitable that the regression line and its tolerance interval have to be extrapolated to achieve a usable result.

Again, it has to be considered whether the treatment of the data should be done strictly according to the rules of statistics, or whether an extrapolation can be allowed. In our view, a slight extrapolation may be possible because the depletion kinetic is assumed to be linear with time ($\log_e$-linearity). Furthermore, tolerance limits are described by hyperbolic curves. Accordingly, the withdrawal period is unlikely to be underestimated when derived by slight extrapolation.

Extrapolation has to be considered with care, when there is indication (e.g. from pharmacokinetic parameters) of a slower final depletion kinetic. Extrapolation far removed from the range of observed data should be avoided. In cases when a withdrawal period can only be derived by a significant extrapolation, further residues data must be provided to confirm the suitability of the derived withdrawal period.

6.3. Should the 95% or the 99% tolerance limit be applied?

Calculations were performed with both the 95% and the 99% one-sided upper tolerance limits (each with a 95% confidence level). Taking into account the MRLs proposed for the target tissues liver and

fat, and using the full data set (including animal 13), withdrawal periods of 28/30 days (95% tolerance limit) and 33/35 days (99% tolerance limit) were calculated. These withdrawal periods were derived by a minimal extrapolation at the 95% tolerance limit for fat and by increased extrapolation at the 99% tolerance limit for both fat and liver.

When applying the 99% tolerance limit one is often confronted with the problem of extreme extrapolation which may result in inadequate withdrawal periods. The 95% tolerance limit in some cases may diminish the extrapolation problem and is therefore expected to provide more realistic withdrawal periods.

For the reasons above the more pragmatic approach - the selection of the 95% tolerance limit for setting withdrawal periods - is preferred.

6.4. Dealing with 'less than' values

Generally, these data cannot be excluded from calculation *a priori*, since they are due to real observations concerning the depletion kinetics. As discussed earlier, setting these data to one-half of the LOD or LOQ should be taken into account. 'Less than' values may also be estimated by special procedures (11, 12).

If, however, the majority of data from one slaughter day are below the LOD (or LOQ) the whole time point should be excluded. This should be the case, especially when the time point in question is a late one which is well off the regression line defined by the other data.

6.5. Dealing with obvious outliers.

For example, could there be any justification to reject the residue data measured for animal 13 of the present data set?

Inspection of the residue data indicated that animal 13 may possibly be an outlier. The residues in all the tissues of this animal (including the injection site) were at or below the LOD at a relatively early time point post dose (day 14, see Table 1). As discussed earlier, the regression assumptions were violated for fat when the full data set was evaluated. Exclusion of animal 13 gave a more reliable basis for the statistical estimation of the withdrawal period.

Usually, due to the limited number of animals and due to the biological animal-to-animal variability, exclusion of values has to be considered with great care. A formal test for outliers has not been recommended in this paper. It may occur, however, that there is a clear reasoning for an exclusion, but removal of data points defined as statistical outliers should only be accepted if there is a **strong causal** justification (e.g. dosing error, sick animals, obvious sampling/analytical error).

6.6. Combining data sets

The benefits and drawbacks of combining studies are discussed in a general section of the 'Guideline on statistical principles for clinical trials for veterinary medicinal products (pharmaceuticals)' (EMA/CVMP/EWP/81976/2010). Generally, such a meta-analysis could have advantages as well as disadvantages: On the one hand, there could be an increase in precision and reliability of results, and sacrificing animals could be reduced. On the other hand, problems might arise if the study characteristics are too different, and if low-quality data are combined with high-quality data, the results might be less reliable than those of an analysis of the high-quality data alone. Thus, combination of data sets might be considered appropriate when the underlying studies are 'similar' and of 'similar quality' (e.g., similar study design, same breeds, animal weight range, dosing, comparable

analytical methods etc.). It would only be appropriate to derive withdrawal periods using the statistical approach, analysing the combined data sets, if the results of the two (or more) studies had been shown to be statistically comparable (for example not statistically different from each other in respect to key parameters such as residual errors of the populations; slope and starting concentrations (C0) of residues. Differences in these and other parameters might indicate differences due to subtle (i.e. not easy to notice) differences in the study designs or other influencing factors.

6.7. The possibility of overriding one study with another

Whether to use or discount a study should depend solely on the quality and validity of the data and not, for example, on the age of the study. Expert judgement is needed, however, to determine whether an 'old' study still reflects contemporary good veterinary and analytical practice (are the animal breeds, treatment and housing conditions and analytical techniques still 'state of the art' and representative of current practices, can these differences have any significant impact on the results?). If old data are considered valid in respect to relevant study design and quality criteria then they should not be discounted in favour of more recently generated residue data.

7. References

1. FDA, General Principles for Evaluating the Safety of Compounds Used in Food-Producing Animals, 1994
2. FDA, General Principles for Evaluating the Safety of Compounds Used in Food-Producing Animals, 1983
3. Lothar Sachs, Angewandte Statistik, 7th Ed., Springer Verlag Berlin, Heidelberg, New York, London, Paris, Tokio, 1992
4. G. Barrie Wetherill, Intermediate Statistical Methods, Chapman and Hall, London, New York, 1981
5. Kurt Stange, Angewandte Statistik, Vol. II, pp. 141-143, Springer Verlag, Berlin, Heidelberg, New York, 1971
6. U. Graf, H.J. Henning, K. Stange, P.T. Wilrich, Formeln und Tabellen der angewandten mathematischen Statistik, 3rd ed., Springer Verlag, Berlin, Heidelberg, New York, London, Paris, Tokio, 1987
7. CVMP, Guideline on the conduct of bioequivalence studies for veterinary medicinal products, EMA/CVMP/016/00-Rev.2, Nov. 2011
8. D.B. Owen, Handbook of Statistical Tables, Addison-Wesley Publishing Company, Reading, Massachusetts, 1962
9. K. Bache, An approach to calculate the inverse of the noncentral t-distribution, to be published, BGA 1993
10. John Mandel, The Statistical Analysis of Experimental Data, Interscience Publ., J. Wiley & Sons, New York, London, Sydney 1964
11. Helsel, D.R., Less than obvious, Envirom. Sci. Technol., Vol 24, No. 12, pp. 1766-1774, 1990
12. Newman, M.C., Dixon, P.M:, Looney, B.B., Pinder, J.E., Estimating mean and variance for environmental samples with below detection limit observations, Water Resources Bulletin, Vol 25, No. 4, pp. 905-916, 1989.

13. CVMP, Guideline on injection site residues, EMEA/CVMP/542/03-FINAL, Apr. 2005

14. CVMP Draft reflection paper on injection site residues: considerations for risk assessment and residue surveillance (EMA/CVMP/520190/2007-Rev.1)

15. VICH GL48: Studies to evaluate the metabolism and residue kinetics of veterinary drugs in food-producing animals: marker residue depletion studies to establish product withdrawal periods(EMA/CVMP/VICH/463199/2009, 14 March 2011)

16. VICH GL49: Studies to evaluate the metabolism and residue kinetics of veterinary drugs in food-producing animals: validation of analytical methods used in residue depletion studies, (EMA/CVMP/VICH/463202/2009, 14 March 2011)

17. Schefferlie & Hekman, The size of the safety span for pre-slaughter withdrawal periods; Journal of Veterinary Pharmacology and Therapeutics Volume 32, Issue Supplement s1, 17 JUL 2009

18. CVMP Draft reflection paper on injection site residues: considerations for risk assessment and residue surveillance (EMA/CVMP/520190/2007-Rev.1).

Annex A

Table 1: Individual results for the marker residue in cattle and calculated daily total residue intake (Data constructed from a real empirical data set)

Animal number	Days post dose	Liver	Fat	Kidney	Muscle	Inj. site	Daily intake*
		(µg/kg)					(µg)
1	7	85.5	96.8	27.0	11.3	123.8	111.0
2	7	141.8	225.0	29.3	11.3	74250.0	37214.7
3	7	198.0	213.8	47.3	15.8	6750.0	3484.5
4	7	31.5	48.3	18.0	4.5	n.a.	-
5	7	119.3	119.3	38.3	9.0	18000.0	9066.0
6	7	108.0	204.8	38.3	18.0	922.5	537.8
7	7	171.0	157.5	6.8	15.8	19125.0	9646.9
8	7	31.5	450.0	11.3	2.3	24.8	99.8
9	7	189.0	65.3	13.5	20.3	4050.0	2101.1
10	7	67.5	195.8	18.0	6.8	495.0	305.6
11	7	135.0	148.5	49.5	20.3	65.3	110.7
12	7	150.8	202.5	60.8	20.3	4500.0	2344.2
13	14	<2.0	<2.0	<2.0	<2.0	2.3	1.8
14	14	22.5	11.3	6.8	2.3	180.0	100.5
15	14	60.8	78.8	20.3	11.3	85.5	79.5
16	14	60.8	51.8	9.0	4.5	2025.0	1042.9
17	14	47.3	33.8	13.5	4.5	121.5	84.4
18	14	22.5	24.8	2.3	2.3	13.5	18.8
19	14	11.3	2.3	2.3	<2.0	<2.0	5.0
20	14	22.5	15.8	13.5	4.5	585.0	304.9
21	14	49.5	51.8	4.5	6.8	49500.0	24775.9
22	14	22.5	13.5	4.5	2.3	105.8	63.6
23	14	40.5	22.5	9.0	4.5	20.3	28.9
24	14	29.3	42.8	18.0	6.8	31.5	35.7
25	21	36.0	27.0	11.3	6.8	33.8	35.3
26	21	9.0	9.0	2.3	2.3	4.5	7.1
27	21	9.0	6.8	2.3	<2.0	<2.0	5.0
28	21	6.8	6.8	2.3	<2.0	<2.0	4.3
29	21	18.0	6.8	2.3	<2.0	<2.0	8.0
30	21	6.8	11.3	2.3	<2.0	<2.0	5.0
31	21	108.0	40.5	11.3	9.0	14850.0	7469.6
32	21	11.3	9.0	4.5	<2.0	11.3	11.7
33	21	2.3	4.5	2.3	<2.0	31.5	17.7
34	21	2.3	9.0	6.8	<2.0	<2.0	3.9
35	21	24.8	9.0	4.5	4.5	11.3	16.2
36	21	2.3	<2.0	<2.0	<2.0	<2.0	1.6
37	28	4.5	4.5	<2.0	<2.0	4.5	4.7
38	28	2.3	4.5	<2.0	<2.0	<2.0	2.2

Animal number	Days post dose	Liver	Fat	Kidney	Muscle	Inj. site	Daily intake*
39	28	11.3	9.0	2.3	<2.0	<2.0	6.2
40	28	9.0	6.8	2.3	<2.0	<2.0	5.0
41	28	<2.0	<2.0	<2.0	<2.0	<2.0	1.2
42	28	4.5	4.5	2.3	<2.0	<2.0	3.1
43	28	<2.0	<2.0	<2.0	<2.0	<2.0	1.2
44	28	<2.0	<2.0	<2.0	<2.0	<2.0	1.2
45	28	2.3	4.5	<2.0	<2.0	<2.0	2.2
46	28	6.8	9.0	2.3	<2.0	<2.0	4.7
47	28	13.5	13.5	4.5	2.0	49.5	32.3
48	28	<2.0	<2.0	<2.0	<2.0	<2.0	1.2
49	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
50	35	n.a.	4.5	n.a.	n.a.	<2.0	-
51	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
52	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
53	35	n.a.	4.5	n.a.	n.a.	4.5	-
54	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
55	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
56	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
57	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
58	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
59	35	n.a.	<2.0	n.a.	n.a.	<2.0	-
60	35	n.a.	<2.0	n.a.	n.a.	<2.0	-

* Amount of total residue calculated by using the ratios marker/total 0.3 for liver, fat, kidney and 0.6 for injection site. The arbitrary food consumption figures used were 100 g liver, 50 g fat, 50 g kidney and 300 g injection site. Values below the limit of detection were set to one-half of the limit of detection (LOD).

n.a.: not assayed

LOD: 2 µg/kg

Results corrected for recoveries

Annex B 1

Comparison to the FDA approach:

In order to compare the results of the equations according to Stange (5) and Graf et al. (6) to the results of the FDA procedure, three data sets out of the data set for liver from Table 1 (Annex A) were tested:

1. The full data set for liver (n=48).
2. The last 5 data of each time point for liver (n=20).
3. The last 3 data of each time point for liver (n=12).

For all three data sets the regression assumptions were met. This can be seen from Table 15.

Table 15: Test results

Data set:	**1**	**2**	**3**
	(n=48)	(n=20)	(n=12)
Bartlett's test	p>0.05	p>0.05	p>0.05
Cochran's test	p>0.05	p>0.05	p>0.05
Lack of fit test	P>0.05	p>0.05	p>0.05
Shapiro-Wilk test	P>0.10	p>0.10	p>0.10

Remark: for all calculation procedures used here values below the LOD were set to one-half of the LOD

Calculation of the tolerance limits:

The tolerance limits according to Stange (5) and Graf et al. (6) were calculated as described earlier (section 2).

The calculation using the non central t-distribution was performed as recommended by the FDA (1, 2):

- calculation of the non-centrality parameter d,
- calculation of the 95th percentile (designated k or t_o of the non-central t-distribution by using the inverse of the noncentral t-distribution function),
- calculation of the tolerance limit according to the equation given in the FDA guideline.

Since the tolerance limits for the calculation of withdrawal periods require only 95% confidence, the tables provided by Owen (8) can also be used. The 95th percentile of the non-central t-distribution for the given non-centrality parameter d and the given degrees of freedom (df=n–2) can be calculated by using the table on page 111 in conjunction with the interpolation procedure described on page 109 of the Owen handbook (8). Because of the very tight tabulation of values the interpolated figures are sufficiently exact. An additional advantage is that the table as well as the interpolation procedure can easily be integrated in any calculation program.

715 Results:

716 1. Data set of 48 animals, 12 per slaughter day, MRL = 30 µg/kg

717 **Table 16:** Upper 95% tolerance limits with 95% confidence

Days post dose	Non-central t-distrib. (µg/kg)	Stange (5) (µg/kg)	Graf et al. (6) (µg/kg)
25	41.60	41.26	41.82
26	36.00	35.70	36.18
27	31.20	30.93	31.35
28	27.07	26.83	27.20
29	23.51	23.30	23.62
30	20.45	20.25	20.53

718 **Table 17:** Upper 99% tolerance limits with 95% confidence

Days post dose	Non-central t-distrib. (µg/kg)	Stange (5) (µg/kg)	Graf et al.(6) (µg/kg)
25	91.20	90.33	92.03
26	78.72	77.94	79.41
27	68.04	67.35	68.62
28	58.88	58.26	59.36
29	51.01	50.46	51.41
30	44.24	43.74	44.57
31	38.40	37.96	38.68
32	33.36	32.96	33.60
33	29.00	28.65	29.20

719 2. Data set of 20 animals, 5 per slaughter day, MRL = 30 µg/kg

720 **Table 18:** Upper 95% tolerance limits with 95% confidence

Days post dose	Non-central t-distrib. (µg/kg)	Stange (5) (µg/kg)	Graf et al.(6) (µg/kg)
25	37.21	36.47	38.00
26	31.98	31.32	32.63
27	27.53	26.95	28.08
28	23.75	23.23	24.21
29	20.52	20.05	20.91
30	17.76	17.33	18.08

721 **Table 19:** Upper 99% tolerance limits with 95% confidence

Days post dose	Non-centr. t-distrib. (µg/kg)	Stange(5) (µg/kg)	Graf et al.(6) (µg/kg)
25	82.57	80.70	85.42
26	70.69	69.02	73.07
27	60.63	59.15	62.63
28	52.10	50.78	53.77

Days post dose	Non-centr. t-distrib. (µg/kg)	Stange(5) (µg/kg)	Graf et al.(6) (µg/kg)
29	44.83	43.66	46.24
30	38.64	37.59	39.83
31	33.35	32.41	34.35
32	28.82	27.98	29.66

722 3. Data set of 12 animals, 3 per slaughter day, MRL = 30 µg/kg

723 **Table 20:** Upper 95% tolerance limits with 95% confidence

Days post dose	Non-centr. t-distrib. (µg/kg)	Stange (5) (µg/kg)	Graf et,al. (6) (µg/kg)
25	88.53	85.10	94.94
26	77.93	74.76	83.45
27	68.79	65.87	73.57
28	60.89	58.19	65.03
29	54.03	51.52	57.63
30	48.04	45.72	51.17
31	42.79	40.64	45.53
32	38.18	36.19	40.58
33	34.12	32.27	36.23
34	30.53	28.82	32.39
35	27.35	25.76	28.99

724 **Table 21:** Upper 99% tolerance limits with 95% confidence

Days post dose	Non-centr. t-distrib. (µg/kg)	Stange (5) (µg/kg)	Graf et al.(6) (µg/kg)
25	240.37	230.00	267.87
26	210.33	200.88	234.02
27	184.56	175.92	205.01
28	162.38	154.44	180.06
29	143.20	135.91	158.52
30	126.57	119.86	139.87
31	112.09	105.91	123.67
32	99.45	93.75	109.54
33	88.39	83.13	97.20
34	78.67	73.83	86.39
35	70.13	65.66	76.89

725
726

Table 22: Withdrawal periods obtained

Data set:	**n=48**		**n=20**		**n=12**	
Tolerance limits*:	95%	99% (days)	95%	99% (days)	95%	99% (days)
Non central t-distribution	28	33**	27	32**	35**	-***
Stange (5)	28	33**	27	32**	34**	-***
Graf et al.(6)	28	33**	27	32**	35**	-***

* with 95% confidence

** more or less severe extrapolation

*** unacceptable extrapolation

Discussion:

Tables 16-21 show that all three methods of calculation gave similar results. When comparing the results of the procedure using the non-central t-distribution to the others, the tolerance limits calculated according to Graf et al (6) were somewhat higher, while those calculated according to Stange (5) were somewhat lower. The time points when the tolerance limits dropped below the MRL of 30 µg/kg are listed in Table 22. As it can be seen in that case, only in one data set (n=12 data set) did a difference of one day appear. The results from Table 22 also show that the evaluation of small data sets (e.g. n=12) could result in relatively long withdrawal periods.

To set withdrawal periods, all three methods of calculation can be considered to be appropriate and of equal value.

With a view to more practical considerations, we propose the procedure according to Stange (6). This approach is not confined to $n \approx 20$, as is the procedure according to Graf et al. (7) and is much easier to perform than the FDA procedure (1, 2) which requires a more elaborate computer program.

Annex B 2

Comparison of different approaches to deal with censored data

In order to compare different approaches to deal with 'less than' values (censored data), the data sets for liver described in Annex B1 were tested by using the following procedures:

- Values below the LOD were excluded (FDA approach)
- Values below the LOD were replaced with LOD/2 (approach currently recommended)
- Values below the LOD were replaced with predicted values (according to the robust method described by Helsel 1990 (11))

Estimated values for the non-detects:

1. Full data set for liver (n=48, see Annex A).
   In the full data set, 1 out of 12 liver samples on day 14 and 4 out of 12 liver samples on day 28 showed values below the LOD (< 2 µg/kg). The predicted values for the non-detects were 10.7 (!) µg/kg for day 14 and 2.0 µg/kg, 1.5 µg/kg, 1.1 g/kg and 0.7 µg/kg for day 28.

As discussed in Section 2 (Step 6) of the main body of this paper, animal 13 is possibly an outlier. This is indicated here by the great difference between the predicted value (10.7 µg/kg) and the observed value (< 2 µg/kg).

2. The last 5 data of each time point for liver (n=20, see Annex A).
   In this data set, only 2 out of 5 liver samples on day 28 yielded values below the LOD. Values of 1.26 µg/kg and 0.46 µg/kg were estimated for these two samples.

3. The last 3 data of each data point for liver (n=12, see Annex A).
   In this data set, the residue concentration of 1 of 3 samples on day 28 was below the LOD. The predicted value for this sample was 3.43 µg/kg.

ad 1. Full data set: 48 animals, 12 per slaughter day:

Table 23: Upper 95% tolerance limits with 95% confidence (non central t-distribution by using the tables provided by Owen (8))

	Values below LOD		
Liver	excluded	LOD/2	predicted values*
Calc.withdrawal period incl. animal 13	27.4***	27.3	25.7
Calc. withdrawal period excl.animal 13**	27.4***	25.7	25.8

* According to Helsel's robust method (11); ** Homogeneity of variances is violated in all three data sets (0.05 >P > 0.025); *** Note that the observed value for animal 13 was a value below the LOD. Consequently, both withdrawal periods are identical.

ad 2. Data set of 20 animals, 5 per slaughter day:

Table 24: Upper 95% tolerance limits with 95% confidence (non central t-distribution by using the tables provided by Owen (8))

	Values below LOD		
Liver	excluded	LOD/2	predicted values*
Calc. withdrawal period	29.6	26.5	26.8

The regression assumption were met in all data sets; * According to Helsel's robust method (11)

ad 3. Data set of 12 animals, 3 per slaughter day:

Table 25: Upper 95% tolerance limits with 95% confidence (non central t-distribution by using the tables provided by Owen (8))

	Values below LOD		
Liver	excluded	LOD/2	predicted values*
Calc. withdrawal period	41.0 ***	34.2**	35.4**

The regression assumption were met in all data sets; * According to Helsel's robust method (11); ** Severe extrapolation; *** Unacceptable extrapolation

The results show that the two substitution methods (i.e. values below the LOD are either replaced with LOD/2 or with the predicted values according to Helsel) resulted in similar withdrawal periods when animal 13 of the full data set (suspected to be an outlier) was excluded from calculation. With the inclusion of animal 13 into the calculation, a shorter withdrawal period was achieved with the Helsel method. This was because the low value of < 2 µg/kg had to be substituted by the high predicted value of 10.7 µg/kg and, therefore, the tolerance interval became closer due to the smaller variance of the data. Omission of the non-detects (FDA approach) resulted in clearly longer withdrawal periods.

Remark: When it is decided to include animal 13 in the calculation, the use of LOD/2 is to be considered rather than the predicted value of 10.7 µg/kg. This is because the value of LOD/2 (1 µg/kg) appears to show more consistency with the observed value (<2 µg/kg).

Annex C

Test of the Significance of the Quadratic Time Effect:

In order to test linearity, checking the significance of the quadratic time effect according to Mandel (10) can be done in advance as an appropriate supplementation to the lack of fit test. The question is, whether a quadratic fit is better than the linear fit.

The linear model is represented by the relation $y = a + bx$, the quadratic model by

$y = a + bx + cx^2$.

Both equations have to be fitted by the method of least squares and the residual errors ($s_{y.x}$) have to be calculated (using the $\log_e$-transformed residue concentrations).

The question is then to determine whether the residual variance of the quadratic fit is significantly smaller than the residual variance of the linear fit. It should be noted, however, that this test only shows if one model is or is not significantly better than the other one, whereas both may be inadequate.

If there is a significant quadratic time effect which is due to the first time point, the next step is to remove the first time point and re-run the analysis.

Remark: A coefficient of the quadratic term equivalent to zero (in the statistical sense) is in accordance with the statement that the linear model is the better one. A statistically significant positive coefficient has to be seen as the most likely alternative model (biphasic elimination kinetic). A statistically significant negative coefficient of the quadratic term indicates that the maximum concentration in tissues has not been reached at early time points.

The test of significance gave the following results for the data for liver and fat from Table 1 (Annex A):

1. Liver

Coefficient c: 0.0017 ± 0.0029 (not significant different from zero at P = 0.05)

Residual error (linear fit): 0.9930

Residual error (quadratic fit): 1.0004

Table 26: Analysis of variance for liver

	Number of parameters in model	**Remaining degrees of freedom**	**Sum of squares of residuals**	**Mean square (SS/df)**
Linear fit:	2	48–2=46	SS_L=45.3569	MS_L= 0.9860
Quadratic fit:	3	48–3=45	SS_Q=45.0339	MS_Q=1.0008
Difference		1	SS_D=0.3230	MS_D= 0.3230

$$\hat{F} = \frac{MS_D}{MS_Q} = \frac{0.3230}{1.0008} = 0.323$$

F (P = 0.05; df1=1, df2=45) = 4.06

Result: The quadratic model is not significantly better than the linear model at the 5% level.

2. Fat:

Coefficient c: 0.0065 ± 0.0029 (not significant different from zero at P = 0.025)

Residual error (linear fit): 1.0258

Residual error (quadratic fit): 0.9839

Table 27: Analysis of variance for fat

	Number of parameters in model	**Remaining degrees of freedom**	**Sum of squares of residuals**	**Mean square (SS/df)**
Linear fit:	2	48–2=46	SS_L=48.4049	MS_L=1.0523
Quadratic fit:	3	48–3=45	SS_Q=43.5584	MS_Q=0.9680
Difference		1	SS_D= 4.8465	MS_D=4.8465

$$\hat{F} = \frac{MS_D}{MS_Q} = \frac{4.8465}{0.9680} = 5.01$$

F (P = 0.05; df1=1, df2=45) = 4.06
F (P = 0.025; df1=1, df2=45) = 5.38

Result: The quadratic model is significantly better than the linear model at the 5% level but not at the 2.5% level. In other words, deviation from linearity emerges.

Conclusion: The quadratic time significance test showed the same results as the lack of fit test (see Step 5 of the draft document). The liver data can be considered linear. For fat, deviation from linearity emerged (0.05 > P > 0.025). As already stated in the main part of the draft document, a re-calculation of the data for fat excluding day 7 from calculation was not taken into account because in our view the linearity assumption was not seriously violated.

Reference:

10. J. Mandel, The Statistical Analysis of experimental Data, Interscience Publ., J. Wiley & Sons, New York 1964.

Annex D

- Compounds for which it was not necessary to establish a MRL (substances with a 'No MRL required' classification):

As stated in the 'Notice to Applicants' for the establishment of MRLs (Volume VIII of the rules governing medicinal products in the EC), a recommendation to insert a compound with status 'No MRL required' in Table 1 of the Annex to Commission regulation (EU) No 37/2010 should not be interpreted as automatically implying that no withdrawal period is necessary.

If there is any indication that the amount of drug derived residues in an edible portion may exceed the ADI, a withdrawal period has to be set. The respective edible portion should include the injection site muscle for substances to be injected intramuscularly or subcutaneously.

Since no MRLs are set for such compounds, the withdrawal period has to be estimated on the basis of the ADI.

For compounds which may cause injection site residues with potential pharmacological effects, it may be necessary to establish a precautionary withdrawal period even when an ADI has not been set (e.g. in the case of hormones the naturally occurring levels in tissues should be used as the starting point for the determination of a withdrawal period). In addition, other reference values may be used, such as daily intake values for vitamins or other food-additives, set by EFSA.

- Generic products:

When the formulation (active and inactive ingredients), the dose schedule, the route(s) of administration and the target species of a specific generic product, are identical to a currently approved product (i.e. the reference product), then the withdrawal period of the latter can be used for the former. However, when there is an indication that the manufacturing process of the generic product may have affected the physicochemical properties of one of the active or inactive ingredients (and in consequence, the bioavailability of the drug), a blood level bioequivalence study is required. This condition, however, only holds true when there is evidence that this modified manufacturing process does not generate impurities or by-products of concern requiring a toxicological re-evaluation.

Demonstration of blood level bioequivalence will also be sufficient to cover differences concerning the formulation of the generic product when the target species and the route of administration are identical.

In the case of products administered subcutaneously or intramuscularly, small differences in composition may have significant effects on injection site depletion which may not be detected in the standard blood level bioequivalence studies. Therefore, for such formulations, in addition to bioequivalence studies, equivalent (or faster) depletion of residue from the injection site should be demonstrated.

In cases where products are intended for administration to the site of action (e.g. topically applied), blood level bioequivalence would not demonstrate the equivalence of local residues. Residues data from the site of administration would be required.

In cases where a change of the target species and/or the route of administration is claimed, information on tissue residue depletion is considered to be necessary. Changes in the dose will also require residue depletion data.

Remark: For experimental design of blood level bioequivalence studies the guideline provided by the CVMP (7) should be taken into account.

Specific problems concerning milk:

See the CVMP Note for guidance for the determination of withdrawal periods for milk (EMEA/CVMP/473/98-FINAL).