CODEX ALIMENTARIUS COMMISSION

Food and Agriculture Organization of the United Nations

World Health Organization

Viale delle Terme di Caracalla, 00153 Rome, Italy - Tel: (+39) 06 57051 - Fax: (+39) 06 5705 4593 - E-mail: codex@fao.org - www.codexalimentarius.net

CX 4/40.2 **CL 2010/24-PR**

TO: Codex Contact Points
Interested International Organizations

FROM: Secretariat, Codex Alimentarius Commission
Joint FAO/WHO Food Standards Programme
FAO, 00153 Rome, Italy

SUBJECT: **Request for comments on the OECD draft calculator and draft user guide**

DEADLINE: **8 August 2010**

COMMENTS:

To:
Ms Barbara MADDEN
Minor Use Officer,
Registration Division,
Office of Pesticide Programs,
United States Environmental Protection Agency
7505p 1200 Pennsylwania Ave, Nw
Washington, Dc 20460
Phone: 703 305 6463
Fax: 703 605 0781
E-mail: madden.barbara@epa.gov

Copies to:
1. Ms Sylvie PORET
Principal administrator
OECD Secretariat
Environment Directorate
2 Rue André Pascal
75016 Paris
France
Phone: 33 1 45 24 89 45
Fax: 33 144 30 63 99
E-mail: Sylvie.Poret@oecd.org

2. Secretariat
Codex Alimentarius Commission
Joint FAO/WHO Food Standards Programme – FAO
Viale delle Terme di Caracalla
00153 Rome, Italy
Fax: +39 (06) 5705 4593
E-mail: codex@fao.org

The last 42nd Session of the Committee on Pesticide Residues had discussed the calculation method for the estimation of maximum residue limits for pesticides being developed through OECD. For details of consideration see ALINORM 10/33/24, paras 134-138. At the above session the Committee agreed to circulate the draft OECD MRL Calculator for comments.

Please note that for testing the calculator you should use the excel file which is attached to the email separately and not the one inside the CL.

Following the decision of the Committee, governments and interested international organizations are invited test the draft Calculator and send any feed-back on both the draft Calculator and draft User Guide as well as the results of any testing they perform, preferably by email, to the above addresses, **before 8 August 2010**.

Compound
Crop
Region / Country
GAP

Reset

Residues (mg/kg)	

Compound
Crop
Region / Country
GAP

Total number of data (n)	0
Percentage of censored data	_
Number of non-censored data	_
Lowest residue (LR)	_
Highest residue (HR)	_
Median residue (STMR)	_
Mean	_
Standard deviation (SD)	_
Correction factor for censoring (CF)	_

Proposed MRL estimate

_

Unrounded	_
Rounded to the next MRL class	_

Rounding system as of 16 March 2010

MRL classes (numerical values)

MRL estimate ≥	Proposed MRL
0.000001	0.01
0.011	0.02
0.021	0.03
0.031	0.04
0.041	0.05
0.051	0.06
0.061	0.07
0.071	0.08
0.081	0.09
0.091	0.1
0.105	0.15
0.155	0.2

0.21	0.3
0.31	0.4
0.41	0.5
0.51	0.6
0.61	0.7
0.71	0.8
0.81	0.9
0.91	1
1.05	1.5
1.55	2
2.1	3
3.1	4
4.1	5
5.1	6
6.1	7
7.1	8
8.1	9
9.1	10
10.5	15
15.5	20
21	30
31	40
41	50
51	60
61	70
71	80
81	90
91	100
105	150
155	200
210	300
310	400
410	500
510	600
610	700
710	800
810	900
910	1000
1050	1500
1550	2000
2100	3000
3100	4000
4100	5000
5100	6000
6100	7000
7100	8000
8100	9000
9100	10000

DRAFT OECD MRL CALCULATOR USER GUIDE
VERSION OF 30TH MARCH 2010

THE OECD MRL CALCULATOR PROJECT

There are two statistically-based calculation procedures in current use around the world for estimation of the MRL/tolerance from supervised field trial data sets: the so-called EU and NAFTA methods. The EU method has now been in use for a number of years in Europe and elsewhere. The NAFTA method, developed by a group of North-American experts, has appeared recently and consequently has not been used as extensively as the EU method. However, both methods have come under criticism (see references below) and some commentators have highlighted apparent shortcomings in both methodologies.

With the goal of addressing those criticisms and harmonizing the calculation of MRLs across the OECD, the OECD Residue Chemistry Expert Group (RCEG) (during its meeting in Washington in 2008) commissioned an expert group formed by regulators and industry specialists to propose a new MRL calculation procedure. The guiding principles of this procedure are:

- the procedure must be a practical implementation of sound statistical methods;
- it must be simple to use without requiring extensive statistical knowledge on the part of the user;
- it should produce a clear and unambiguous MRL proposal for most residue datasets produced by field trials; and,
- it should harmonise the EU and NAFTA procedures as much as possible.

Following these guiding principles, the OECD RCEG MRL Calculation Group is currently working on the development and implementation of a robust methodology which seems to produce satisfactory results for the considerable number of real residue datasets tested so far. The latest version of this procedure is distributed jointly with this document.

The statistical goal of the OECD MRL calculator, in common with previous methodologies, is to produce an MRL proposal in the region of the 95th percentile of the underlying residue distribution (which we abbreviate as p95), which is conservative in the sense that it will have a much greater propensity to make errors by overestimating p95 than by underestimating it for most datasets.

This distribution of the calculator and the user guide has the objective of obtaining feedback from the WGP, RSG, RCEG and Ad-hoc Technical Policy Advisory Group on the MRL Calculator, and it should not be distributed further nor should it be considered to be the final project output.

HOW TO USE THE OECD MRL DRAFT CALCULATOR SPREADSHEET

To compute an MRL, the user inputs the data in the left-most blue column of the "Input-Output" sheet under "Residues (mg/kg)". Censored data (residue values that are less than the limit of quantification or LOQ) are entered by listing the LOQ value (example, 0.01) along with an asterisk in the adjacent column. The order in which the data are entered does not impact the results. For residue tests with replicate field samples, the average or mean of the replicate values is used for the respective test in the calculator. The spreadsheet then automatically conducts all the computations and reports all the relevant results on the same page. Above the column for the residue data, four cells with text fields are available for documenting the dataset.

To go back to a clean spreadsheet, the user may click on the "Reset" button.

EXPLANATORY MESSAGES DISPLAYED BY THE CALCULATOR

If the dataset consists of less than three values the message "MRL calculation not possible (too small dataset)" is displayed at the bottom of the spreadsheet. The choice of 3 values was made based on the minimal requirement common among OECD countries. In theory, it is possible to perform an MRL calculation with only two values but the level of uncertainty of the proposal would be ridiculously high.

With a single residue value, it is impossible to compute an estimator for the standard deviation of the dataset, which is needed in the calculation procedure.

If the dataset consist of less than eight residue values, a warning message is printed before the displayed number of residue measurements indicating the considerable level of uncertainty surrounding the calculation of any statistical quantity for such small datasets. For a dataset with eight residue values, the estimated failure rate, i.e. the probability that the MRL is below the 95th percentile of the residue distribution, reaches approximately 25 %.

Similarly, a warning message is displayed if more than 50% of the dataset is censored (residues below the limit of quantification or LOQ1) for the same reason. Although the methods selected for the MRL calculation are very robust to the presence of non-detected residues, uncertainty is considerable for residue datasets for which the majority of residues values are LOQs.

HOW THE OECD MRL DRAFT CALCULATOR WORKS

The results of the computation are displayed in a frame at the right of the input data. The first few fields provide some general information about the dataset, including values for the highest residue (HR) and the supervised trial median residues (STMR). Please, notice the range of the residue values by comparing the lowest residue with the HR; the larger that range is, the greater the variability present in the data. This variability is taken into account when computing the MRL.

The algorithm for the computation of the MRL is very simple. The maximum of three calculated results is put forward as the MRL proposal by the calculator:

1. the mean and the standard deviation values of the dataset are computed; the "mean + 4* standard deviation" value is evaluated as the base proposal;

2. the "3*mean" value is also computed to provide a "floor" to the calculation; that is to guarantee that the sample coefficient of variance (CV = standard deviation / mean) used in the calculation is at least 0.5 (a correction factor has recently being added because it was observed that the mean of a dataset is overestimated for censored datasets; the correction factor CF is equal to 1 – ⅔*fraction censored data in the dataset[2]);

3. the HR value is also used as a "floor" to guarantee than the MRL proposal is always greater than or equal to the highest residue.

So the MRL proposal will be determined by the formula:

Maximum (mean + 4*standard deviation, 3*mean*CF, HR).

Finally a comment about datasets that are fully or almost fully censored: For fully censored datasets, the selected procedure will produce an MRL equal to the highest LOQ present in the dataset. For small datasets (less than twenty LOQs), this proposal should also be supported with some experimental evidence that residues are not expected for this particular use, since statistically there may still be a chance of getting a residue value above the highest LOQs. For datasets mostly made of a single LOQ but with a few actual measurements at that LOQ level, the above procedure (after rounding) will automatically produce an MRL higher than that LOQ level. The case of almost fully censored datasets but with several LOQ levels is more complicated and should be considered on a case-by-case basis.

ROUNDING

To facilitate the use of MRLs in the regulations, MRL proposals are rounded as a last step in the calculation. For numbers between 1 and 10, we round to a single digit; for 10 to 100, we round to units of 10; for 100 to 1000 we round to units of 100 and so on. Intermediate values of 0.15, 1.5, 15, etc, were introduced to avoid

[1] No distinction is made between the level of quantification (LOQ) and the level of detection (LOD). For all non-quantified residue measurements, it is the LOQ that should be input in the calculation.

[2] This is our most recent modification of the calculation procedure and it is still under evaluation; as such, it is particularly subject to potential changes.

doubling of MRLs on rounding (except for the lowest order of magnitude considered by the calculator for which this would not be practical). So for example: 0.12 rounds up to 0.15, 0.16 rounds up to 0.2; and 12 rounds up to 15 instead of 20. The possibility for rounding down exists if a particular MRL level is surpassed by a small amount.

To be more precise, the rounding possibilities are (in mg/kg)[3]:

0.01		0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	1.5	2	3	4	5	6	7	8	9
10	15	20	30	40	50	60	70	80	90
100	150	200	300	400	500	600	700	800	900
1000	...								

MRLs should be displayed without zeroes to the right of the left most significant digit, to avoid giving the impression of having more accuracy than what exists in reality. So, for example, an MRL should be displayed as 2 mg/kg but not 2.0 mg/kg; 0.1 mg/kg is possible but 0.10 mg/kg is not.

Rounding down will happen if the MRL proposal does not exceed the lower MRL rounding possibility by more than 10% of the difference between the lower and the next MRL rounding possibilities. As a result, rounding down will happen if the right-most digit to the most significant one that is going to be kept is zero (0). The exception is between 1 and 2. Some examples of cut off points:

MRL Class	**10% of Difference**	**Cut off Point**
0.01	0.001	0.011
0.02	0.001	0.021
0.03	0.001	0.031
...	...	....
0.09	0.001	0.091
0.1	0.005	0.105
0.15	0.05	0.155
0.2	0.01	0.21
0.3	0.01	0.31
...	...	...
0.9	0.01	0.91
1	0.05	1.05
1.5	0.05	1.55
2	0.1	2.1
3	0.1	3.1

Some rounding examples:

[3] The smallest rounding possibility of 0.01 mg/kg was set from feedback obtained from a number of regulatory authorities that completed the policy questionnaire circulated last year through relevant OECD and JMPR committees.

Unrounded proposal:	0.0108 mg/kg	→	Rounded proposal: MRL:	0.01 mg/kg
Unrounded proposal:	0.0112 mg/kg	→	Rounded proposal: MRL:	0.02 mg/kg
Unrounded proposal:	1.04 mg/kg	→	Rounded proposal: MRL:	1 mg/kg
Unrounded proposal:	1.12 mg/kg	→	Rounded proposal: MRL:	1.5 mg/kg
Unrounded proposal:	1.53 mg/kg	→	Rounded proposal: MRL:	1.5 mg/kg
Unrounded proposal:	1.58 mg/kg	→	Rounded proposal: MRL:	2 mg/kg
Unrounded proposal:	2.07 mg/kg	→	Rounded proposal: MRL:	2 mg/kg
Unrounded proposal:	2.12 mg /kg	→	Rounded proposal: MRL:	3 mg/kg
Unrounded proposal:	21.0 mg/kg	→	Rounded proposal MRL:	20 mg/kg

STATISTICAL DISCUSSION

The procedure described above was chosen by the calculator group because it is very robust and simple to use, having shown in our extensive simulations to outperform more sophisticated methods for the dataset sizes usually available for the computation of MRLs around the world (see calculations below). The performance of the procedure was tested on both synthetic and real datasets. Also, MRL proposals were compared with historical MRLs of EFSA, JMPR, and NAFTA.

The "Mean + 4*SD" base method was chosen for its independence of distributional assumptions4, its robustness in relation to the presence of censored data (LOQs) and its good performance for small datasets compared to the distributional methods. To help correct for possible underestimations of the standard deviation for very small datasets, a minimal CV requirement of 0.5 was established which produced the "floor" method "3*Mean."5 The further requirement that the MRL proposal should be greater or equal to the highest residue (HR) was introduced from feedback obtained through the circulation for over a year in various JMPR and OECD committees of a questionnaire that contained policy questions needed to complete the design of the MRL calculator.

All three requirements (the base method "Mean + 4*SD" and the requirements of a CV of 0.5 and that the MRL should be higher or equal than the HR) plus the correction factor for censoring, are compactly expressed by the formula:

Maximum (mean + 4*standard deviation, 3*mean*CF, HR).

PERFORMANCE AGAINST SYNTHETIC DATA

Testing on synthetic datasets was performed with 100,000 datasets sampled from the lognormal distribution with the CV = 1.06. This distribution was believed to represent a reasonable worst case for real field trial data, which means that performance is expected to be better than depicted below for most datasets.

For each dataset generated from the lognormal distribution, a MRL proposal was calculated. For the smallest dataset that we consider (with at least 3 data points), most of the calculated MRL proposals (95%) lay between 0.37*p95 and 4.5*p95 (Fig. 1), while the MRL-over-HR ratio varied between 2.0 and 2.7 (Fig. 2). We call these intervals the 95% probable ranges of the computation, because 95% of the time the results are within that range. The failure rate, i.e. the chance to get a MRL below the p95, was about 42.2% for 3 data points (Fig. 3). It decreased to approximately 25% for 8 data points, and the level of 5% was reached

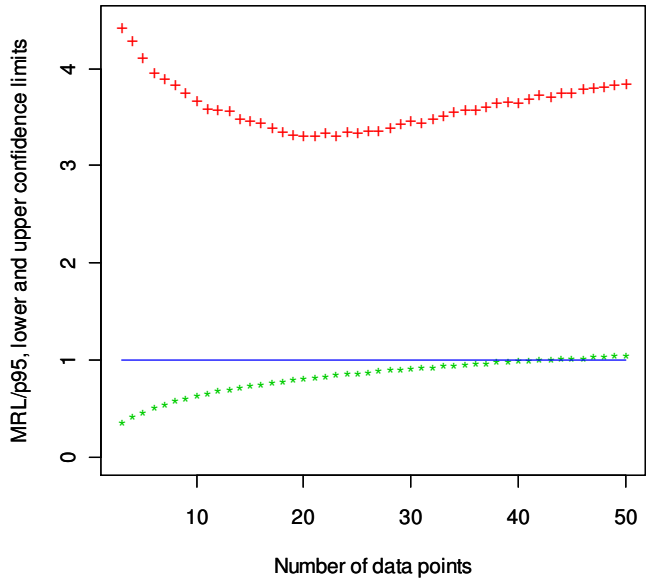
[4] Although this method does not depends on any distributional assumption, a classical theorem called the Chebishev's inequality states that for any large enough sample extracted from any distribution with finite mean and variance, the "Mean + 4*SD" method will provide an estimate for a percentile above the 93th percentile. For the lognormal distribution, it will provide estimates above the 99th percentile for large enough samples.

[5] If the CV = 0.5, then "SD = 0.5*Mean" and then we have that "Mean + 4*SD" = "Mean + 2*Mean" = "3*Mean."

[6] For the underlying normal distribution, the mean was 1 and the standard deviation was $\ln(2)^{1/2} \sim 0.83$.

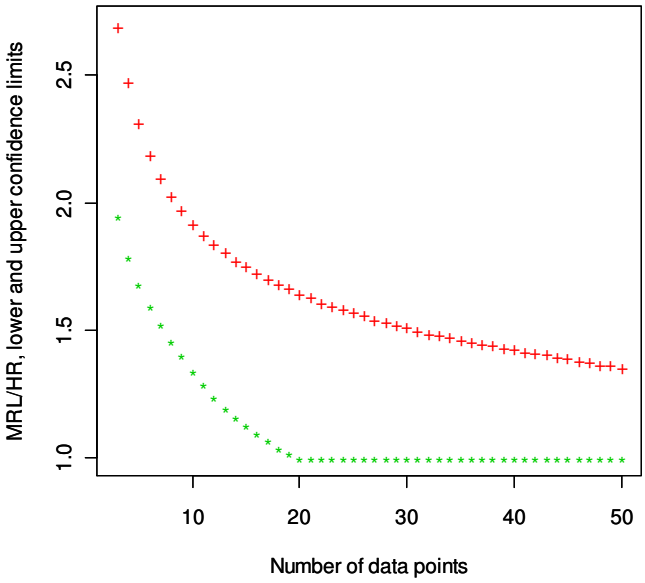
for 29 data points.

Figure 1. Figure 1. 95% probable ranges for the MRL-over-p95 ratio depending on the number of data points



Red and green labels show the upper and lower boundaries, respectively. The blue line represents the equality of the ratio.

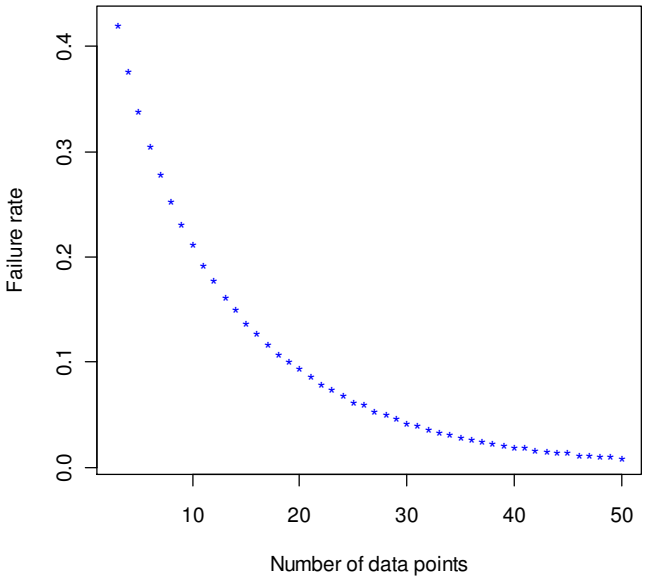
Figure 2. 95% probable ranges of the MRL-over-HR ratio depending on the number of data points.



. Red and green plus signs show the upper and lower boundaries, respectively

Figure 3. "Failure rate"

i.e. a fraction of datasets for which the proposed MRL is below the p95, versus the number of data points.



So even for this reasonable worst case and for an extremely small datasets, the MRL proposal is expected to be roughly between half and four times the 95th percentile. Even for the extremely small datasets, there is a greater probability that the MRL proposal will be above the 95th percentile instead of below it (almost 60% confidence).

PERFORMANCE AGAINST REAL DATA (SUB-SAMPLING FROM LARGE DATASETS)

A considerable amount of effort and time was dedicated to compare the performance of the different calculation options considered during the course of this project against real datasets. In this user guide we will only report the results for the suggested method.

For a real dataset, the underlying distribution as well as any of its percentiles is unknown. To evaluate the performance against real datasets of any methodology that was under consideration, we selected large well-behaved real datasets, we extracted multiple subsets of different sizes from them, and we computed MRL proposals for all these subsets.7 The HR of the large dataset is expected to be in the range of the highest percentiles of the underlying distribution (typically above the 95th percentile for a dataset of more than 20 residue values); so it can be taken as a reference point as is done in Figure 4.

As can be seen in that figure, for subsets of 16 and 20 points, most of the MRL proposals are above the HR of the full dataset. For datasets of 4 and 8 points, most MRL proposals lie between 0.5 and 2.5 of the HR of the parent dataset. Even for these small subsets, the MRL proposal is above the HR of the parent dataset much more frequently than below.

Another way of evaluating at the performance (and also the robustness) of the calculator is displayed in Figure 5. The ratio of the MRL proposal for the subset over the full dataset is displayed. For all dataset sizes, there is a very clear peak around 1, which means a perfect match. For very small datasets, the peak ratios level off, producing values between 0.5 and 1.5.

[7] We sample without replacement, so this is similar to bootstrapping but not identical.

Figure 4. Ratio of the MRL proposal for each subset and the HR of the full set.

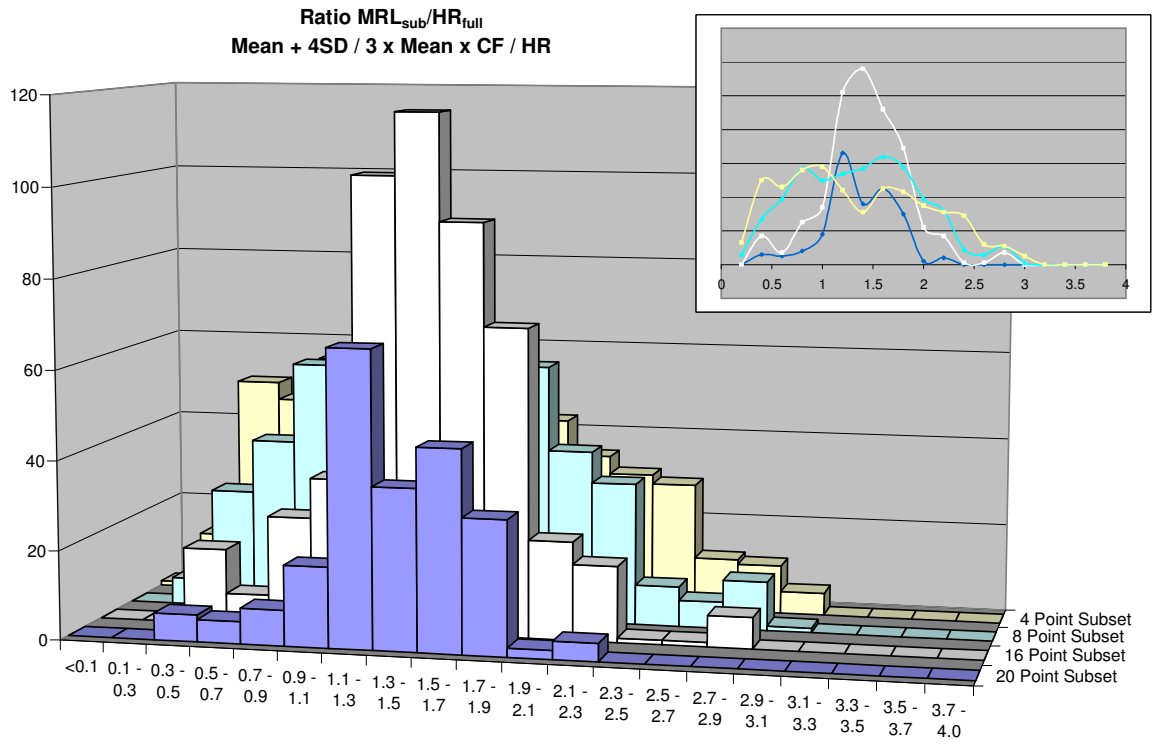


Figure 5. Ratio of the MRL proposal for the subset and the MRL proposal for the full set.

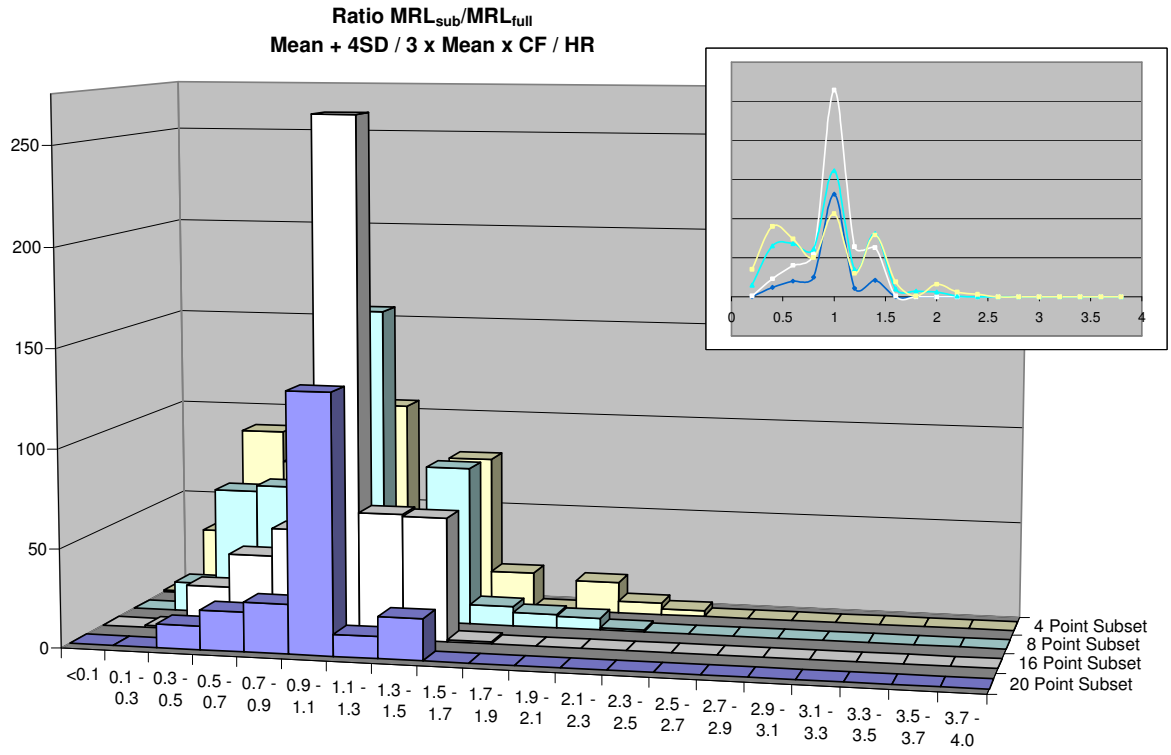
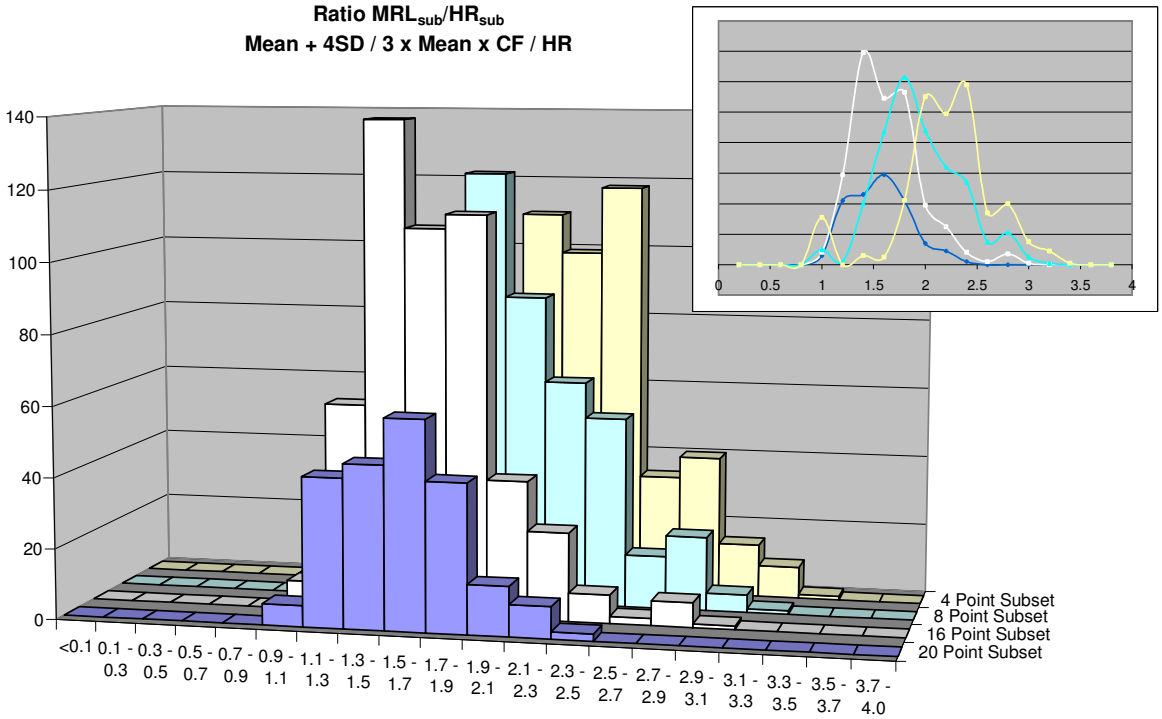


Figure 6. Ratio of the MRL proposal of a subset to its HR.



Finally in Figure 6, the ratio of the MRL proposal of a subset to its HR is shown to illustrate that for very small datasets, it may even be above 3.

COMPARISON WITH HISTORICAL DATASETS

The proposed method – maximum (mean + 4*standard deviation, 3*mean*CF, HR) – was tested using real residue data that were recently evaluated by experts of EFSA or JMPR and used by these experts to derive MRLs. The residue data evaluated by EFSA were taken from "Reasoned Opinions on MRLs" published by EFSA between November 2008 and February 2010 (ref. 5). The residue data evaluated by JMPR were taken from the JMPR Report 2008 (ref. 6). Some of the characteristics of the two residue data collections are shown below:

Table 1. Characteristics of EFSA and JMPR residue data collections.

	EFSA	JMPR
Total number of datasets	215	201
Number of different active substances	47	15
Number of different commodities	Ca. 70	ca. 80
Number of datasets with $n \leq 4$	52	17
Number of datasets with $4 < n \leq 8$	110	79
Number of datasets with $8 < n \leq 16$	44	62
Number of datasets with $n > 16$	9	43
Number of datasets without any censored value	148	120
Number of fully censored datasets	14	15

In case of real datasets, where the underlying distributions are not known, it is difficult to assess whether the MRLs proposed by the calculator are suitable or not. Thus the performance of the calculator was characterized by doing some statistics on:

- the ratio between the rounded MRL and the HR;
- the ratio between the rounded MRL and the unrounded MRL;
- the ratio between the rounded MRL and the EU MRL proposed by EFSA or JMPR; and,
- the ratio between the rounded MRL and the MRL derived using the NAFTA calculator (note: when using the NAFTA calculator censored data were not replaced by maximum likelihood estimates).

For each of these ratios the minimum, maximum and mean were determined for all the EFSA and all the JMPR datasets. Furthermore the mean was also calculated by grouping the datasets according to size and percentage of censored data (please see to table 2 and table 3).

Table 2. Test with the proposed MRL calculator using EFSA residue data.

EFSA datasets	Rounded MRL / Unrounded MRL	Rounded MRL / HR	Rounded MRL / EFSA MRL	Rounded MRL / NAFTA calc. MRL
Min (overall)	0.95	1.00	0.40	0.43
Max (overall)	4.16	10.31	2.00	2.67
Mean (overall)	1.19	2.05	1.12	1.11
Mean ($n \leq 4$)	1.30	2.54	1.08	1.16
Mean ($4 < n \leq 8$)	1.19	2.03	1.12	1.09
Mean ($8 < n \leq 16$)	1.12	1.68	1.21	1.08
Mean ($n > 16$)	1.07	1.35	0.92	1.08
Mean (0% censored)	1.11	2.03	1.16	1.12
Mean (100% censored)	1.03	1.03	0.91	1.00

Table 3. Tests with the proposed MRL calculator using JMPR residue data.

JMPR datasets	Rounded MRL / Unrounded MRL	Rounded MRL / HR	Rounded MRL / JMPR MRL	Rounded MRL / NAFTA calc. MRL
Min (overall)	0.96	1.00	0.40	0.50
Max (overall)	1.82	3.13	3.00	3.00
Mean (overall)	1.10	1.79	1.05	1.16
Mean ($n \leq 4$)	1.09	2.35	1.28	1.37
Mean ($4 < n \leq 8$)	1.08	1.96	1.08	1.12
Mean ($8 < n \leq 16$)	1.10	1.63	1.00	1.12
Mean ($n > 16$)	1.15	1.49	0.97	1.22
Mean (0% censored)	1.09	1.95	1.05	1.16
Mean (100% censored)	1.00	1.00	1.04	1.00

On average the rounding procedure tends to increase the MRLs by about 10—20%. With some EFSA datasets, a more than 3-fold increase due to rounding was observed. This is because the LOQ for these datasets was 0.00097 mg/kg and the unrounded MRLs were far below the lowest MRL class of 0.01 mg/kg.

The mean ratio between the rounded MRL and the HR was 2.1 and 1.8 for the EFSA and JMPR dataset collections, respectively. This ratio tends to decrease when the size of the dataset increases. A more than 10-fold ratio was observed for some of the EFSA datasets that were analysed with an LOQ of 0.00097 mg/kg and which contained no data higher than the LOQ.

On average the MRL estimates yielded by the draft calculator exceed the MRLs proposed by EFSA and JMPR experts by 12% and 5%, respectively. However, larger deviations are observed for some individual datasets. The ratio between the MRL estimates produced by the draft calculator and the MRLs proposed by experts ranges between 0.40 and 2.0 for the EFSA datasets and between 0.40 and 3.0 for the JMPR datasets.

On average the MRL estimates yielded by the draft calculator also tend to exceed the MRLs yielded by the NAFTA calculator. The average exceedance is 11% for the EFSA datasets and 16% for the JMPR datasets. Again, larger deviations are observed for some individual datasets. The ratio between the MRL estimates produced by the draft calculator and the MRLs produced by the NAFTA calculator ranges between 0.43 and 2.7 for the EFSA datasets and between 0.50 and 3.0 for the JMPR datasets.

The following graphs allow for comparison of the MRLs produced by the draft calculator (Y-axis) with the MRLs proposed by EFSA or JMPR experts as well as with the MRLs produced by the NAFTA calculator (X-axis). Both axes are represented using a logarithmic scale. The points on the blue line correspond to datasets for which the draft calculator yields an MRL-estimate that is equal to the MRL proposed by experts or produced by the NAFTA calculator. Points above (below) the line correspond to datasets for which the draft calculator yields an MRL-estimate that is higher (lower) than the MRL proposed by experts or produced by the NAFTA calculator.

Figure 7. Comparison between the MRLs produced by the draft calculator and the MRLs proposed by EFSA experts

(EFSA datasets).

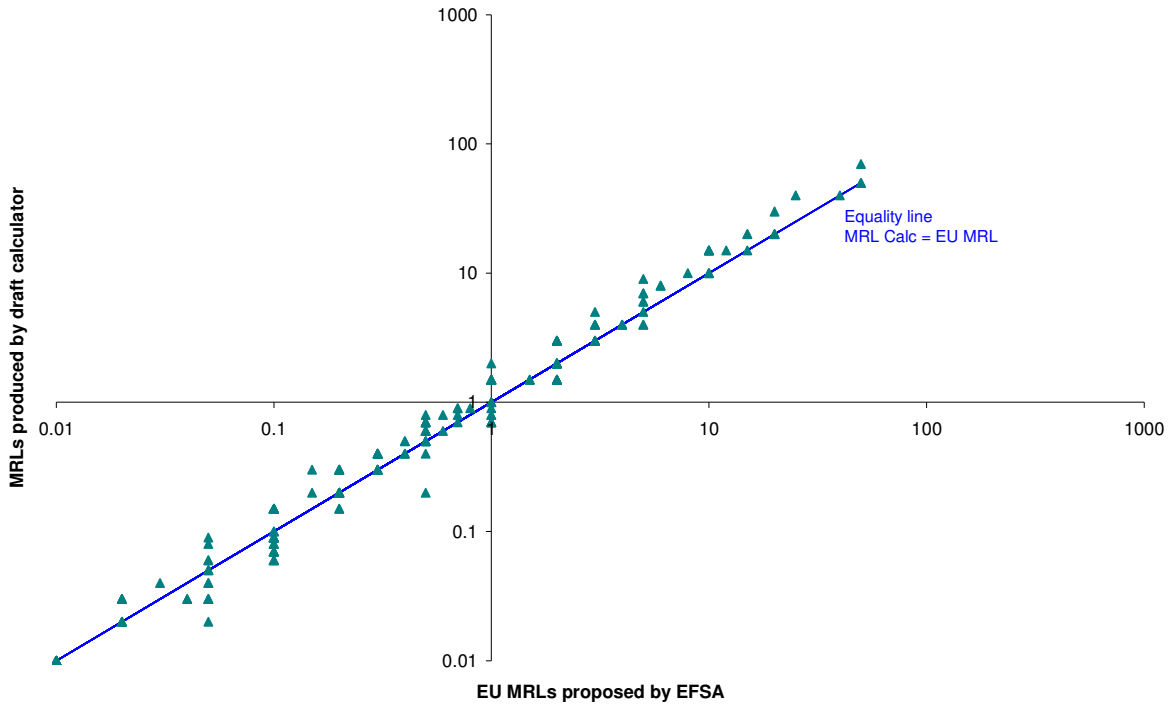


Figure 8. Comparison between the MRLs produced by the draft calculator and the MRLs produced by the JMPR experts

(JMPR datasets).

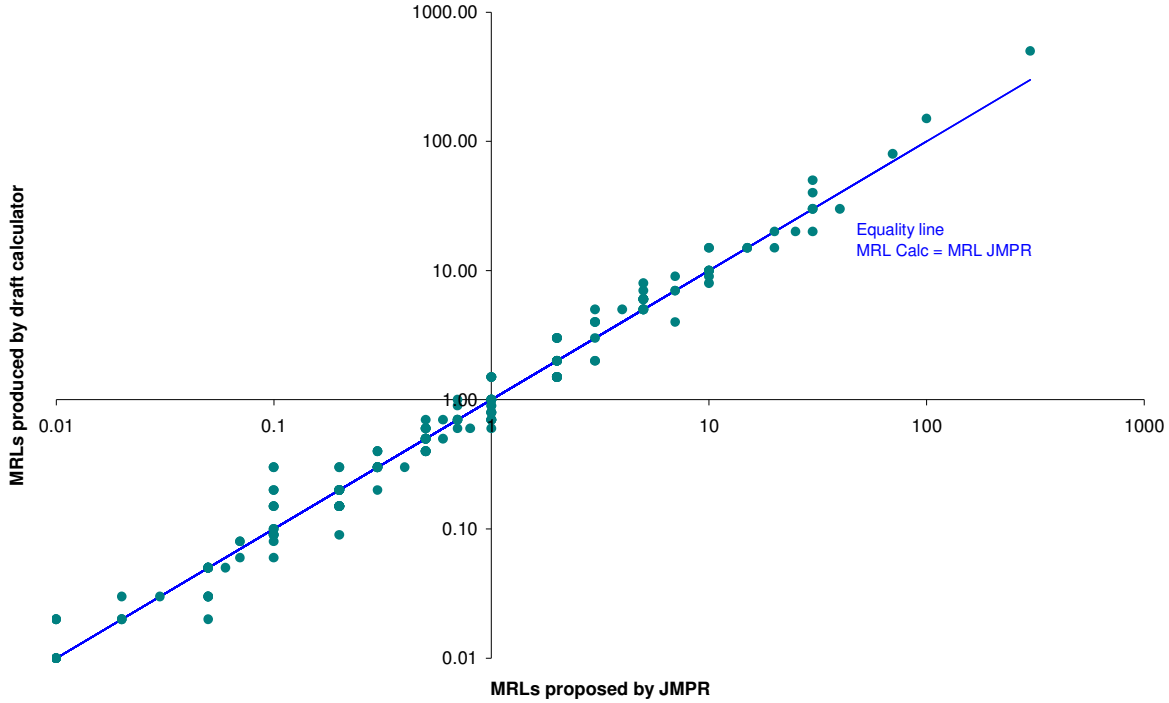
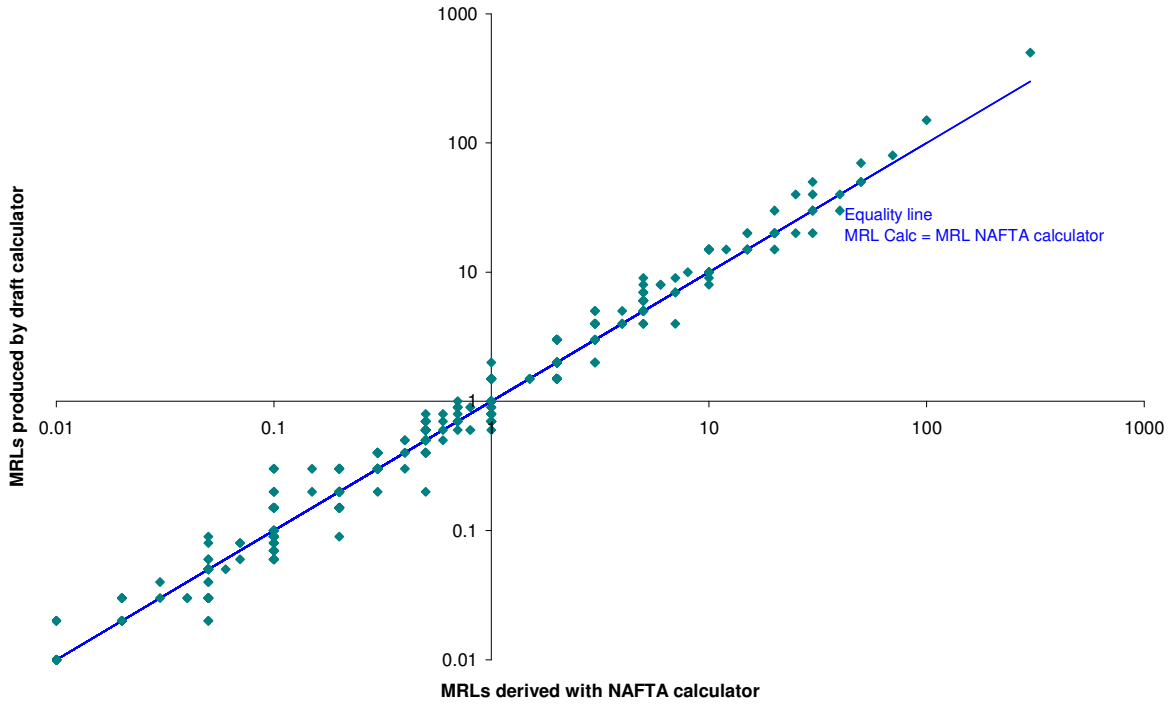


Figure 9. Comparison between the MRLs produced by the draft calculator and the MRLs produced by the NAFTA calculator

(combined EFSA and JMPR datasets).



ALTERNATIVE METHODS

The recommended methodology was tested against a number of alternative methods. Only the most relevant contenders are listed below. More information will be included in the statistical white paper that will be produced and distributed by the group in due course (current deadline: July 2010).

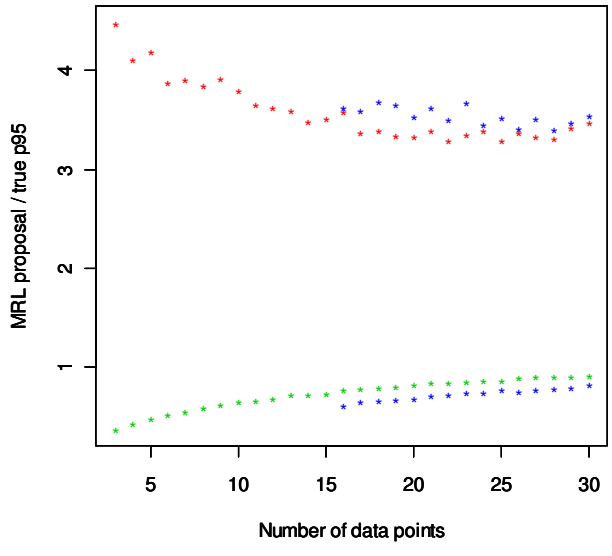
COMPARISON WITH DISTRIBUTIONAL METHODS

The original design of the calculator was based on the expectation that distributional approaches (fitting to normal, lognormal and Weibull distributions) would be more accurate than non-distributional approaches for large datasets. However, the calculator work group has found that the "Mean + 4*SD" method outperformed the distributional approaches, not just for small datasets, but also for datasets as large as 20 or 30 points.

To compare the performance of the proposed method with the previously used distributional method, 10,000 datasets were sampled from the lognormal, normal and Weibull distributions with CV = 1.0 for each dataset size from 3 to 30.[8] For each distribution and for each number of data points, the 95% probable range[9] of the ratio of the calculated MRL to the true p95 was calculated. Negative values sampled from the normal distribution were replaced with 0.001.

Figures 9—11 demonstrate that in general the proposed method outperformed the distributional ones for dataset sizes from 16 to 30 (for smaller datasets the distributional method was not applied). Indeed, for lognormal and Weibull datasets (Fig. 9 and 11, respectively), the probable range was narrower, and both the upper and lower boundaries of the MRL-over-p95 ratio were closer to 1.0 compared to the distributional method, i.e. the 95th percentile of the residue distribution was estimated more accurately. For normal datasets the distributional method provided the MRL-over-p95 closer to 1.0 and narrower probable range (Fig. 10). On the other hand, the proposed method performed more conservatively; the entire probable range was above the 95th percentile for datasets of 16 and more points. The distributional method, in contrast, underestimated the 95th percentile of the underlying normal distribution even for medium size datasets.

Figure 9. Lognormal datasets: 95 % probable range for the MRL-over-p95 ratio depending on the number of data points.

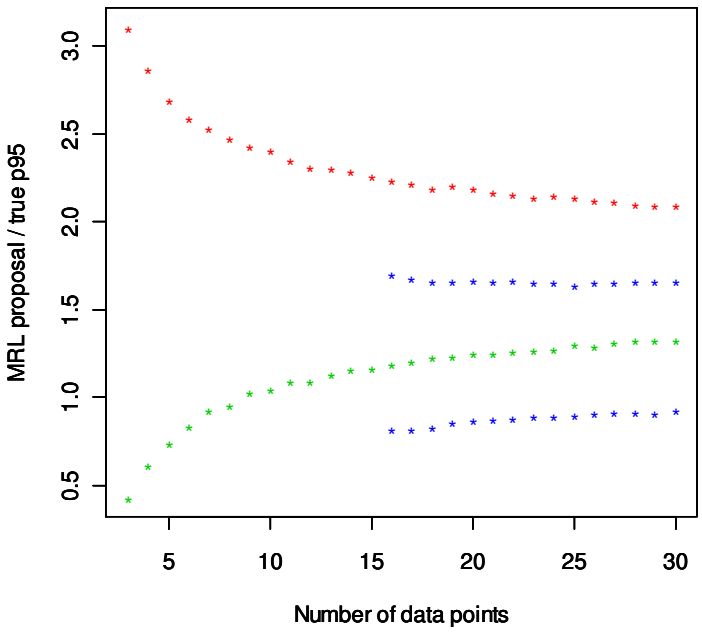


Red and green labels show the upper and lower boundaries, respectively, calculated by the proposed method. Blue labels show values calculated by the distributional method.

[8] Lognormal distribution with the mean of logs = 1.0 and SD of logs = 0.83, normal distribution with the mean = 1.0 and SD = 1.0 and Weibull distribution with the shape parameter = 1.0 and scale parameter = 1.0 were considered.

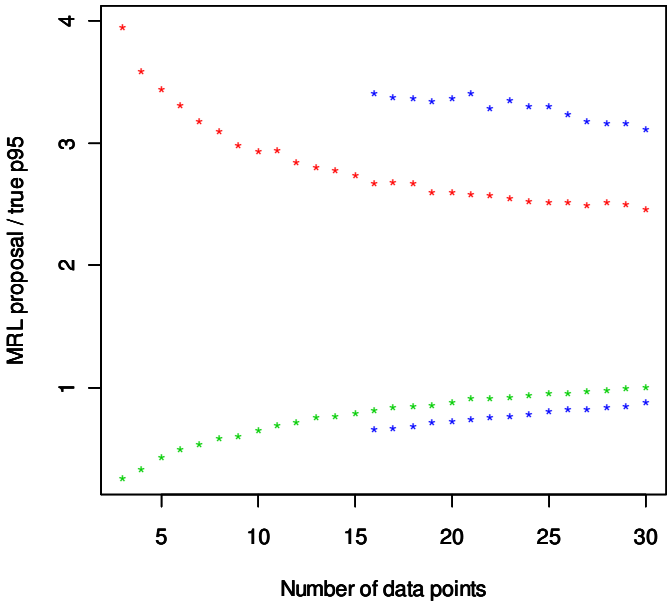
[9] The interval between 2.5% lowest value and the 97.5% highest value of the ratio computed from all the datasets.

Figure 10. Normal datasets: 95 % probable range for the MRL-over-p95 ratio depending on the number of data points.



Red and green labels show the upper and lower boundaries, respectively, calculated by the proposed method. Blue labels show values calculated by the distributional method.

Figure 11. Weibull datasets: 95 % probable range for the MRL-over-p95 ratio depending on the number of data points.



Red and green labels show the upper and lower boundaries, respectively, calculated by the proposed method. Blue labels show values calculated by the distributional method.

COMPARISON WITH MODIFIED VERSIONS OF THE SUGGESTED PROCEDURE

As an alternative to introducing a “floor” method to reduce underestimations of the 95th percentile for very small datasets, a number of other procedures were investigated. Most of these procedures adjusted the

number of standard deviations used in the calculation according to dataset size.

By the "Mean + k*SD" method we mean that k is an integer factor that depends on the number of data points n, so we have k = k(n). If the value of "Mean + k(n)*SD" is less than the highest residue (HR), then the HR is taken as an MRL proposal. So the MRL proposal is

Maximum (mean + k(n) * SD, HR).

Several criteria were suggested to calculate the k factors, i.e. the functional dependency of k on the number of data points n. Below we report the results of on just one of them; the other criteria produced similar results.

Let the target criteria be that MRL proposals are 1) higher than p95, in line with the EU and NAFTA recommendations, and 2) lower than F*p95 (F is an integer factor) with a high level of confidence q. The default level of confidence was set to q = 95 %, in line with the EU and NAFTA recommendations. However, for small datasets, simultaneous satisfaction of requirements 1) and 2) turn out to be impossible. In such cases the level of confidence was decreased as a function of dataset size. Factor F was set to 4 to provide a) not too high MRL estimates and b) reasonable levels of confidence for small datasets.

To derive the dependencies k(n) and q(n), 100,000 synthetic datasets sampled from the lognormal distribution with the CV of 1.0 were considered. For each dataset in each data size class, a MRL proposal was calculated. A k value was accepted for a given number of data points if upper and lower confidence limits of MRL proposals were below 4*p95 and above 1.0*p95, respectively (requirements 1) and 2)). If necessary, the confidence level q was decreased from 95 %.

Calculated values of k varied from 9 for 3 data points to 4 for 39 or more data points (Fig. 12). The level of confidence was in the worst case (3 data points) 60 %, while the level of 95 % was reached for 22 data points (Fig. 13). In the worst case (3 data points) most of the calculated MRL proposals (95 %) lay between 0.4*p95 and 9.0*p95 (Fig. 14), while the MRL/HR ratio varied between 1.9 and 5.1 (Fig. 15). On the other hand, if MRL proposals were considered at calculated levels of confidence, they would be, by the definition of the procedure, within required upper and lower boundaries for all dataset sizes (Fig. 14, blue labels). The failure rate, i.e. the chance to get a MRL below the p95, was at maximum 19.1 % for n = 3 data points (Fig. 16), and fast decreased to a level of about 2—3% with an increase of n.

Figure 12. Calculated k values depending on the number of data points.

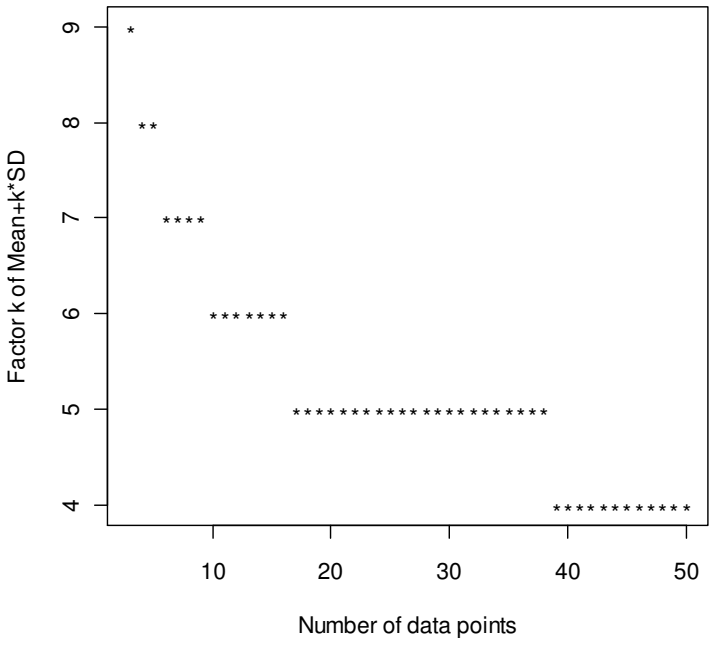


Figure 13. Calculated levels of confidence depending on the number of data points.

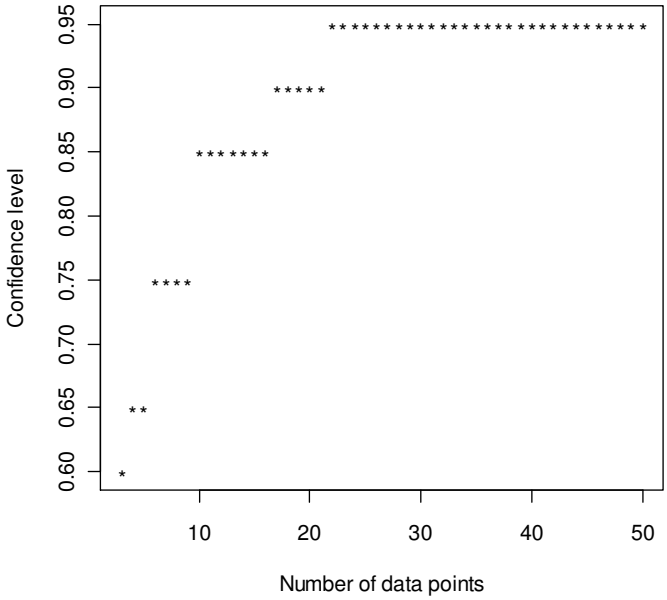
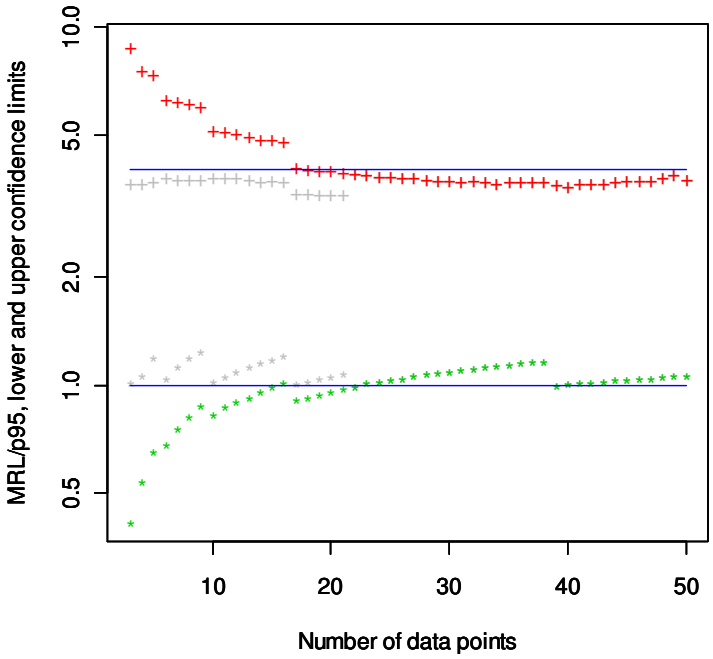
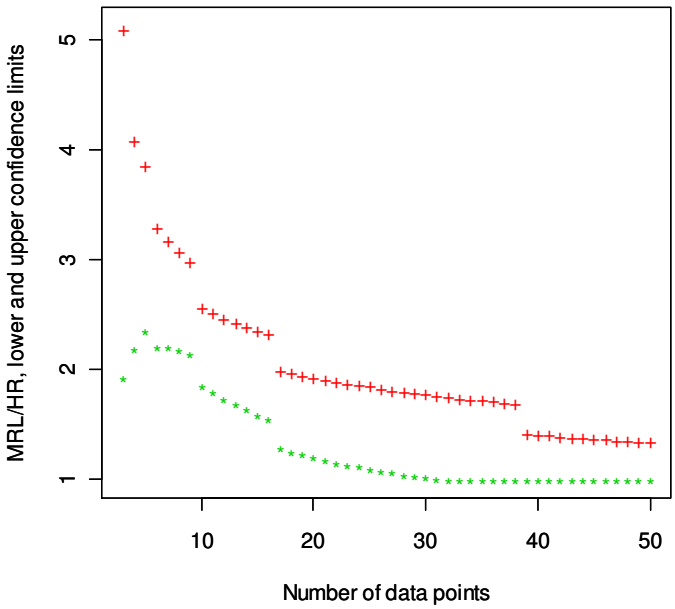


Figure 14. 95% probable ranges for the MRL-over-p95 ratio depending on the number of data points.



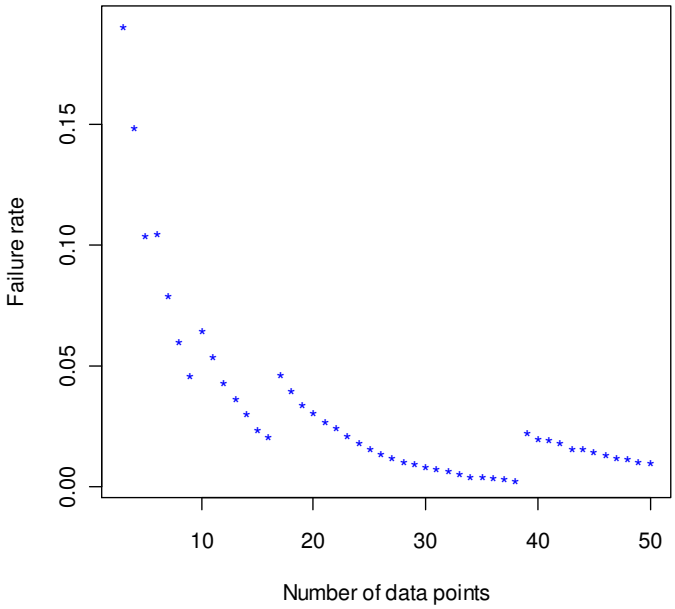
Red and green labels show the upper and lower boundaries, respectively. Blue labels show values calculated using 1*p95 and 4*p95 as lower and upper limits for MRL proposals.

Figure 15. 95% probable range for the MRL-over-HR ratio.



Red and green labels show the upper and lower boundaries, respectively.

Figure 16. "Failure rate", i.e. the fraction of datasets for which the proposed MRL is below the p95, versus the number of data points.



The "Mean + k*SD" method was tested using real residue data from EFSA and JMPR evaluations. The test results are summarised in table 4 and table 5.

Table 4. Tests with the "Mean + k*SD" method using EFSA residue data.

EFSA datasets	Rounded MRL / Unrounded MRL	Rounded MRL / HR	Rounded MRL / EFSA MRL	Rounded MRL / NAFTA calc. MRL
Min (overall)	0.91	1.00	0.40	0.67
Max (overall)	10.31	10.31	3.00	3.75
Mean (overall)	1.19	2.86	1.59	1.53
Mean ($n \leq 4$)	1.33	3.45	1.51	1.55
Mean ($4 < n \leq 8$)	1.16	2.94	1.65	1.59
Mean ($8 < n \leq 16$)	1.12	2.26	1.63	1.42
Mean ($n > 16$)	1.06	1.43	0.97	1.14
Mean (0% censored)	1.12	2.94	1.68	1.58
Mean (100% censored)	1.03	1.03	0.91	1.00

Table 5. Tests with the "Mean + k*SD" method using JMPR residue data.

JMPR datasets	Rounded MRL / Unrounded MRL	Rounded MRL / HR	Rounded MRL / JMPR MRL	Rounded MRL / NAFTA calc. MRL
Min (overall)	0.95	1.00	0.50	0.63
Max (overall)	1.42	5.00	5.00	2.86
Mean (overall)	1.09	2.37	1.38	1.47
Mean ($n \leq 4$)	1.08	3.12	1.74	1.70
Mean ($4 < n \leq 8$)	1.09	2.82	1.56	1.56
Mean ($8 < n \leq 16$)	1.07	2.08	1.26	1.41
Mean ($n > 16$)	1.12	1.65	1.06	1.32
Mean (0% censored)	1.09	2.62	1.41	1.51
Mean (100% censored)	1.00	1.00	1.04	1.00

On average the rounding procedure tends to increase the MRLs by about 10—20%. With some EFSA datasets, a more than 10-fold increase due to rounding was observed. This is because the LOQ for these datasets was 0.00097 mg/kg and the unrounded MRLs were far below the lowest MRL class of 0.01 mg/kg.

The mean ratio between the rounded MRL and the HR was 2.9 and 2.4 for the EFSA and JMPR dataset collections, respectively. This ratio tends to decrease when the size of the dataset increases. A more than 10-fold ratio was observed for some of the EFSA datasets that were analysed with an LOQ of 0.00097 mg/kg and which contained no data higher than the LOQ.

On average the MRL estimates yielded by the draft calculator exceed the MRLs proposed by EFSA and JMPR experts by 59% and 39%, respectively. However, larger deviations are observed for some individual datasets. The ratio between the MRL estimates produced by the draft calculator and the MRLs proposed by experts ranges between 0.40 and 3.0 for the EFSA datasets and between 0.50 and 5.0 for the JMPR datasets.

On average the MRL estimates yielded by the draft calculator also tend to exceed the MRLs yielded by the NAFTA calculator. The average difference is 53% for the EFSA datasets and 47% for the JMPR datasets. Again, larger deviations are observed for some individual datasets. The ratio between the MRL estimates produced by the draft calculator and the MRLs produced by the NAFTA calculator ranges between 0.67 and

3.8 for the EFSA datasets and between 0.63 and 2.9 for the JMPR datasets.

The following graphs allow us to compare the MRLs produced by the "Mean + k*SD" method (Y-axis) with the MRLs proposed by EFSA or JMPR experts as well as with the MRLs produced by the NAFTA calculator (X-axis). Most points are located above the blue line, which indicates that the tested "Mean + k*SD" method tends to produce MRLs that are both higher than the MRLs recently proposed by EFSA and JMPR experts and higher than the MRLs produced by the NAFTA calculator.

Figure 17. Comparison between the MRLs produced by the "Mean + k*SD" method and the MRLs proposed by EFSA experts

(EFSA datasets).

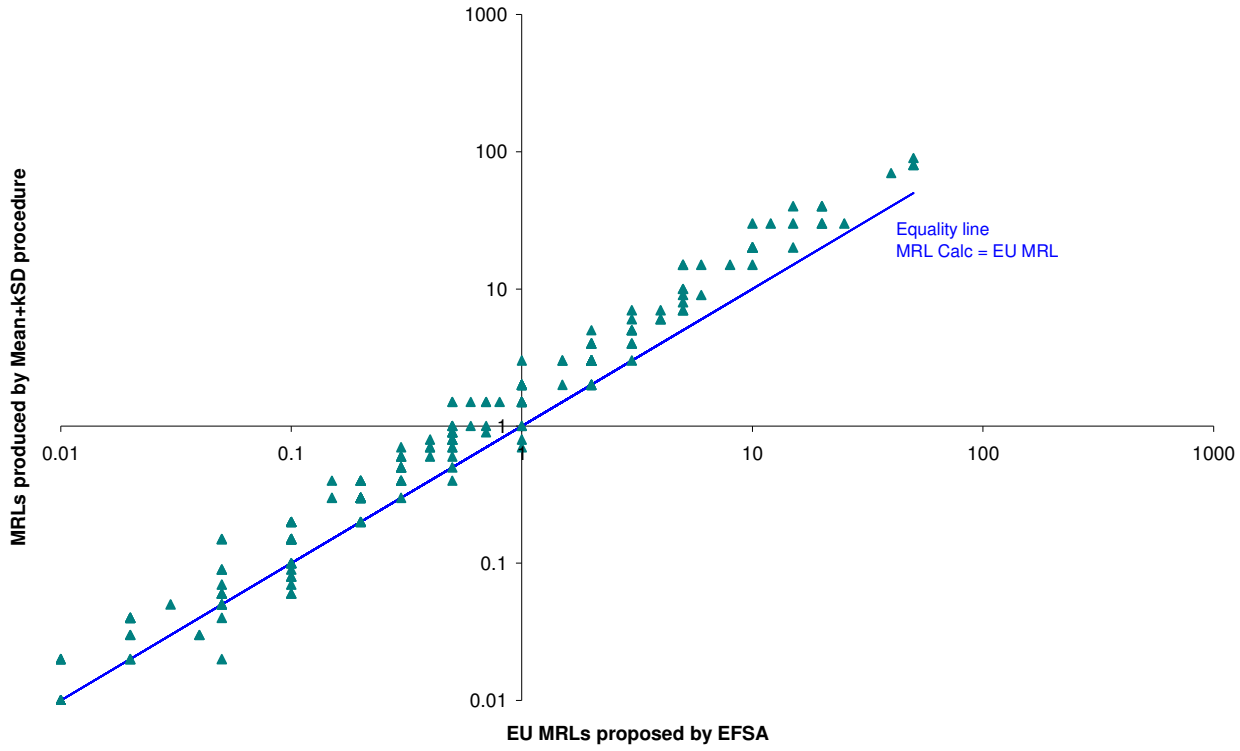


Figure 18. Comparison between the MRLs produced by the "Mean + k*SD" method and the MRLs proposed by JMPR experts

(JMPR datasets).

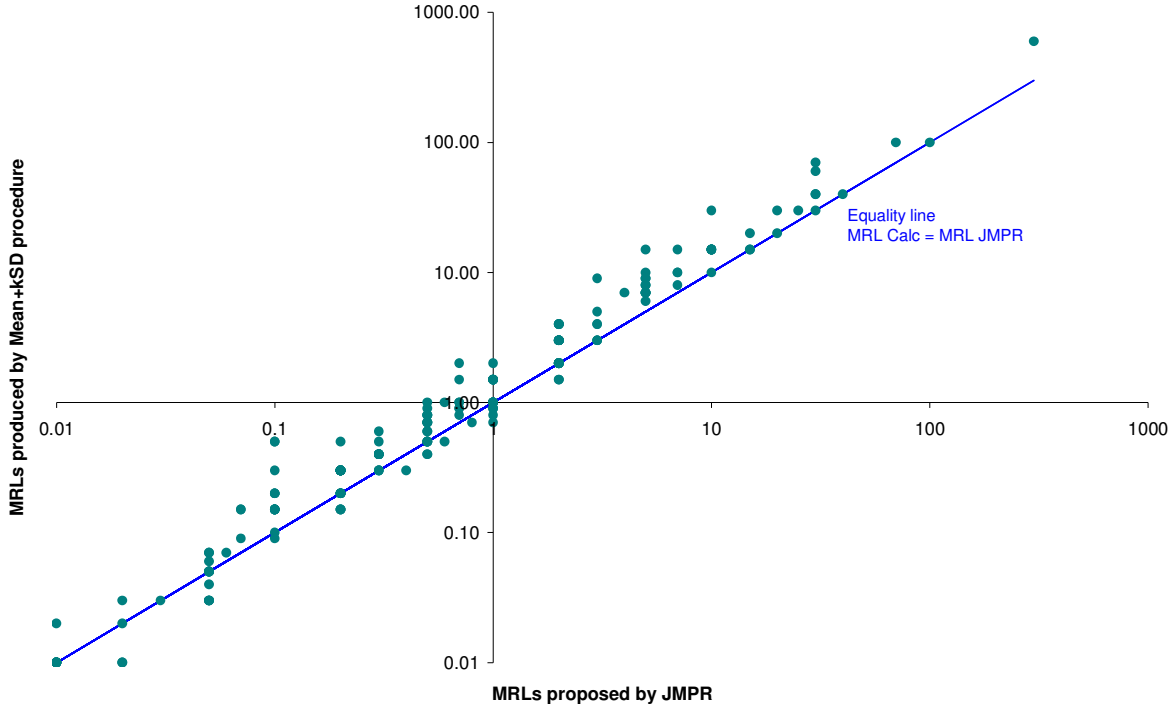
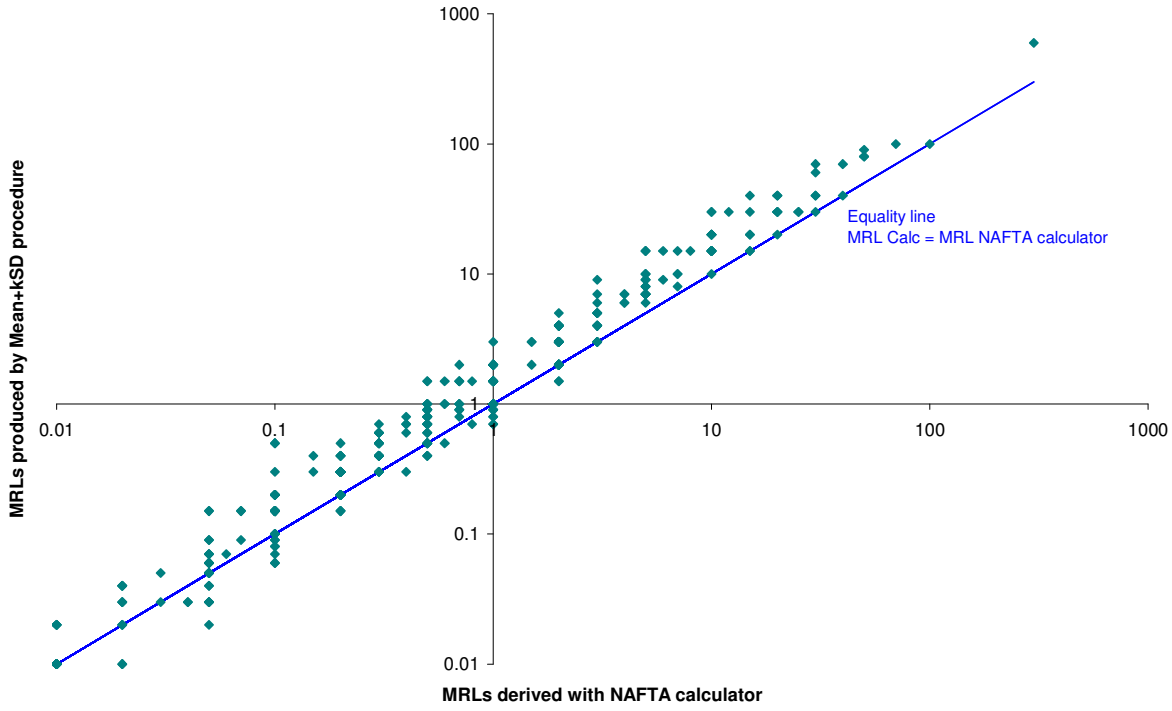


Figure 19. Comparison between the MRLs produced by the "Mean + k*SD" method and the MRLs produced by the NAFTA calculator (combined EFSA and JMPR datasets).



In general, this method is more conservative, not only more than the one suggested by the MRL calculation group, but also more than any other MRL calculation procedures ever used before. Of course, the selection of the target criteria for any MRL calculator is a policy decision that corresponds to the risk managers.

CONCLUSIONS

All the procedures included in this section are scientifically sound, being based in well established statistical procedures. They differ on their degree of conservatism and their performance with small datasets.

The suggested base method, "Mean + 4*SD", was expected to be more robust and outperform distributional methods for very small datasets. Indeed a similar method to this one was already introduced as part of the NAFTA calculator. But it was a pleasant surprise to find out that it also outperforms or at least, performs similarly than the distributional methods for medium-size datasets of 20 to 30 points.

Our attempts at reducing the underestimation of the 95th percentile for very small datasets by making the number of standard deviations dependent on the dataset size produce much more conservative methods and increase overestimations. The addition of a floor method was a pragmatic decision born from this experimentation and it is oriented to correct the worst underestimations without increasing the overestimations.

In relation to the overestimations, the issue of potential "outliers" was discussed many times by the group. It was decided that it is very difficult to even determine if a certain high value should be called an outlier for small datasets, since there is not enough information in them to determine the "trend" of the dataset. For larger datasets, which have been shown to clearly follow a certain distribution (normal, lognormal, Weibull, etc), an appropriate outlier test for that distribution may be performed; e.g., the Dixon test on the residues directly for the normal distribution, or in the logarithms of the residues for the lognormal distribution. This may highlight certain residues measurements that deserve further investigation. But the group would not recommend ignoring these data points exclusively based on the results of statistical tests.

REFERENCES

Maximum Residue Levels: Fact or Fiction? Kieran Hyder et al. Human and Ecological Risk Assessment, 9(3):721-740, 2003.

Statistical Basis of the NAFTA Method for Calculating Pesticide Maximum Residue Limits from Field Trial Data. Available from the EPA website.

Review of the EU, NAFTA and CODEX Procedures for MRLs Calculation. José Domingo Salazar. European Crop Protection Association, 2007.

Comparing Two Methodologies for Setting MRLs from Field Trial Data. DSI Consulting Pty Ltd, 2000.

These 69 documents may be downloaded on the EFSA website :http://www.efsa.europa.eu/en/scdocs.htm by searching under the Unit "PRAPeR" and the document type "reasoned opinion".